University of Alberta

Characterization of a Cyanobacterial RNA Helicase Gene

by

Wendy Colleen Magee

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science

in

Microbiology and Biotechnology

Department of Biological Sciences

Edmonton, Alberta

Fall 1997

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-22631-X

Canada

University of Alberta

Library Release Form

Name of Author: Wendy Colleen Magee

Title of Thesis: Characterization of a Cyanobacterial RNA Helicase Gene

Degree: Master of Science

Year This Degree Granted: 1997

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly, or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Wendy Magee

3242-104A Street
Edmonton, Alberta
T6J 2Z6

August 18, 1997

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Characterization of a Cyanobacterial RNA Helicase Gene** submitted by **Wendy Colleen Magee** in partial fulfillment of the requirements for the degree of **Master of Science** in **Microbiology and Biotechnology**.

Dr. George W. Owttrim

Dr. Julia M. Foght

Dr. Brenda K. Leskiw

Dr. John B. Bell

Aug. 9/97
Date

DEDICATION

To my parents, David and Bernice, who have supported
me in the completion of this study.

ABSTRACT

A partial putative RNA helicase gene, designated gwo29, was used to screen *Anabaena variabilis* UTCC 387 genomic clone banks for the presence of the full-length gene. Three clones were identified by this method, two of which had identical insert sizes of 4 kb. The third was not stable in *Escherichia coli* and was lost. One of the two identical clones, designated pWM75, was characterized further using a number of approaches. Restriction endonuclease mapping was performed and sequencing of a 1964 bp internal region revealed an open reading frame of 1278 bp. These data showed that the deduced amino acid sequence of this open reading frame contains all of the highly conserved amino acid motifs found in DEAD box RNA helicases. The gene was thus designated *crhA* for *c*yanobacterial *R*NA *h*elicase *A*. The CrhA protein was overexpressed in *E. coli* using the pGEX system. However, numerous attempts to purify the CrhA protein for use both in enzyme assays (ATP and RNA binding, RNA-dependent ATPase, and ATP-dependent RNA unwinding) and in the production of antibodies were unsuccessful. *In vivo* inactivation of the *crhA* open reading frame by insertional inactivation resulted in the generation of merodiploids that had no obvious aberrant phenotype on solid medium under normal growth conditions but that had an altered morphology in liquid under these same conditions. The results are discussed in terms of the study of RNA helicases in general and of their significance to *A. variabilis* UTCC 387 in particular.

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to the following people who have made the completion of this study possible:

Dr. G. Owttrim for his supervision, advice and guidance.

My supervisory committee, Dr. B. Leskiw and Dr. J. Foght for their helpful suggestions and encouragement.

Sonya Kujat for her moral support, valuable help and for being a good listener.

Dr. D. Chamot for her advice, support and for proofreading this thesis.

Dr. W. Page and Dr. L. Frost for the use of lab equipment and products.

Rich Mah and Jeff Hoyem for coming to the rescue every time lab equipment or the computer broke down.

Randy Mandryk and the University of Alberta Photographic Services for taking all of the photographs in this thesis.

Bill Klimke for his assistance with computer sequence analyses.

The friends I have made over the past two years, in addition to the ones mentioned above, who have made this a fun place to work, especially Christie Hamilton, Richard Fekete, Anne Sharpe, Esther Yu, Tony Cornish and Wally Blank.

My brother Shawn, for typing Chapter Four.

Finally, I cannot say thanks enough to Bev Aindow, who not only typed Chapters One through Three but who also did all of the schematic diagrams and the photograph labelling - it looks great!

TABLE OF CONTENTS

		Page Number
	Abstract	
	Acknowledgements	
	List of Tables	
	List of Figures	
	List of Abbreviations	
	CHAPTER 1 - INTRODUCTION	1
1.1	RNA Helicases:	1
1.1.1	General Characteristics of RNA Helicases	1
1.1.2	eIF4A	2
1.1.3	RNA Helicase Amino Acid Sequence Homology	4
1.1.4	RNA Helicase Involvement in Cellular Processes	9
1.1.5	Prokaryotic RNA Helicases	11
1.2	Cyanobacteria:	16
1.2.1	General Characteristics of Cyanobacteria	16
1.2.2	The Genus *Anabaena*	18
1.2.3	Genetic Systems in *Anabaena*	18
1.2.4	*Anabaena* sp. Strain PCC 7120	21
1.2.5	Use of *Anabaena* sp. Strain PCC 7120 as a Model System	26
1.3	Thesis Objectives	26
	CHAPTER 2 - MATERIALS AND METHODS	28
2.1	Bacterial Strains and Growth Conditions	28
2.2	Purification of DNA:	32
2.2.1	Small Scale Purification of *E. coli* Plasmid DNA	32
2.2.2	Large Scale Purification of *E. coli* Plasmid DNA	32
2.2.3	Purification of *A. variabilis* UTCC 387 Chromosomal DNA	32
2.2.4	Purification of *E. coli* Chromosomal DNA	34
2.3	Manipulation of DNA:	35
2.3.1	Digestion and Gel Electrophoresis of DNA	35
2.3.2	Purification of DNA From Agarose Gels	36

2.3.3	Southern Transfer and Hybridization	37
2.3.4	*In situ* Hybridization of Bacterial Clones	38
2.3.5	DNA Sequencing	40
2.3.6	DNA Ligation	43
2.3.7	Bacterial Transformation	43
2.3.8	Bacterial Electroporation	44
2.4	Overexpression of Protein in *E. coli*:	45
2.4.1	Cloning of the *chrA* Open Reading Frame into Expression Vector pGEX-2T	45
2.4.2	Batch Purification of Fusion Protein	46
2.5	Protein Manipulation:	48
2.5.1	SDS-Polyacrylamide Gel Electrophoresis of Proteins	49
2.5.2	Staining Proteins in Polyacrylamide Gels	50
2.6	Cyanobacterial Gene Inactivation:	50
2.6.1	*In vitro* Inactivation of *chrA*	50
2.6.2	*In vivo* Inactivation of *chrA*	51
	CHAPTER 3 - RESULTS	53
3.1	Isolation of an RNA Helicase From *A. variabilis* UTCC 387	53
3.2	Mapping of pWM75	56
3.3	Sequencing of the pWM75 Insert and Sequence Analysis	60
3.4	Overexpression of CrhA	74
3.5	Inactivation of *crhA*	82
	CHAPTER 4 - DISCUSSION	93
	Bibliography	105
	Appendix	117

LIST OF TABLES

Table Number		Page Number
2.1	Bacterial Strains	29
2.2	Parent Plasmids	30
2.3	Plasmid Constructions	31
2.4	Oligonucleotide Primers	41
3.1	Comparison of Percent Identity and Percent Similarity Among Selected RNA Helicase Amino Acid Sequences	73

LIST OF FIGURES

Figure Number		Page Number
1.1	Schematic Diagram of DEAD Box RNA Helicases	5
1.2	Schematic Representations of the DEAH and DEXH Families of RNA Helicases	7
3.1	*A. variabilis* UTCC 387 *Xba*I Genomic Clone Banks	55
3.2	Southern Hybridization of *A. variabilis* UTCC 387 and *E. coli* DH5α Genomic DNAs with pWM75	57
3.3	Restriction Map and Sequencing Strategy of pWM75	58
3.4	Southern Analysis of pWM75	61
3.5	Nucleotide Sequence of the *Hinc*II/*Eco*RV Region of the pWM75 Insert	63
3.6	Alignment of RNA Helicase Amino Acid Sequences	67
3.7	Flowsheet of Site Specific Mutagenesis and Isolation of the DNA fragment Encoding the *crhA* Open Reading Frame	75
3.8	Overexpression of CrhA in *E. coli*	78
3.9	Construction of the *crhA* Inactivation Plasmid	83
3.10	Morphology of Wild Type and Mutant *A. variabilis* UTCC 387 Cells	87
3.11	Southern Analysis of Spectinomycin/Streptomycin Resistant *A. variabiliis* UTCC 387 Genomic DNA	89
3.12	*In vivo* Inactivation of *crhA*	91

LIST OF ABBREVIATIONS

APS	ammonium persulfate
ATP	adenosine 5'-triphosphate
bp	basepair(s)
BSA	bovine serum albumin
°C	degrees centigrade
cm	centimeter(s)
cm^2	centimeter(s) square
cpm	counts per minute
CTAB	hexadecyltrimethylammonium bromide
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	any deoxynucleoside triphosphate
DTT	dithiothreitol
EDTA	ethylene diamine tetra acetic acid
g	gram(s)
h	hour(s)
IPTG	isopropyl - β - D - thiogalactoside
kb	kilobase(s)/kilobasepair(s)
kDa	kiloDalton(s)
L	liter(s)
M	molar

m^2	meter(s) square
mg	milligram(s)
min	minute(s)
mL	milliliter(s)
mM	millimolar
MOPS	3-[N-Morpholino]propanesulfonic acid
N	normal
nm	nanometer(s)
Ω	ohm(s)
OD	optical density
%	percent
pmol	picomole(s)
PVP	polyvinylpolypyrrolidone
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
sarkosyl	*N*-lauroylsarcosine
SDS	sodium dodecyl sulfate
sec	second(s)
SSC	sodium chloride/sodium citrate
TAE	Tris/acetate/EDTA
TBE	Tris/borate/EDTA

TEMED	N,N,N',N' - tetramethylethylenediamine
Tris	Tris (hydroxymethyl) aminomethane
U	unit(s)
μF	microfarad(s)
μg	micrograms(s)
μL	microliter(s)
UV	ultraviolet
V	volt(s)
v/v	volume per volume
W	watt(s)
w/v	weight per volume
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

CHAPTER ONE
INTRODUCTION

1.1 RNA HELICASES

1.1.1 General Characteristics of RNA Helicases

The modulation of RNA structure is a critical step in many fundamental cellular processes including translation, ribosome biogenesis and mRNA splicing (Fuller-Pace, 1994). RNA helicases play a major role in these processes through their ability to change RNA conformation. These enzymes possess known or putative ATP-dependent RNA unwinding activity (Fuller-Pace, 1994) as well as RNA-dependent ATPase activity (Gorbalenya and Koonin, 1993). In combination, these activities allow RNA helicases to utilize the energy obtained from ATP hydrolysis in order to unwind duplex regions of RNA, whether this be double-stranded RNA or base-paired regions in single-stranded RNA.

As a result of this enzymatic activity, RNA helicases can be placed within a larger group of DNA- and/or RNA-dependent nucleoside triphosphate phosphatases possessing helicase activity (Gorbalenya, *et al.*, 1989). Based on amino acid sequence comparisons, this broad group of proteins has been divided into three large superfamilies and two smaller families (Gorbalenya and Koonin, 1993). RNA helicases are members of superfamily SF2, a group which also includes a number of DNA repair, recombination and replication enzymes (Gorbalenya and Koonin, 1993).

Amino acid sequence comparisons are utilized not only to define families of helicases, but also to identify individual members since enzymatic activity has been demonstrated in only a few cases. Definitive assignment of a protein as an RNA helicase

therefore requires not only sequence homology but biochemical evidence as well.

1.1.2 eIF4A

The prototype RNA helicase is eIF4A (*e*ukaryotic translation *i*nitiation *f*actor *4A*), an enzyme required for ribosome binding to mRNA during translation initiation in eukaryotes (Rozen, *et al.*, 1989). eIF4A is a protein of 50 kDa which exists as both a free form and as part of eIF4F, a multisubunit complex consisting of eIF4E (the cap binding protein), eIF4G (a protein required for the structural integrity of the complex), and eIF4A (Rozen, *et al.*, 1990).

eIF4A possesses a number of activities consistent with those of an RNA helicase. This protein was shown to exhibit ATP-dependent RNA binding (Grifo, *et al.*, 1984) with no sequence specificity and a preference for single-stranded RNA (Abramson, *et al.*, 1987). ATP hydrolysis was required for the reaction as shown by the inability of a non-hydrolyzable analogue of ATP to promote binding (Grifo, *et al.*, 1984). eIF4A also displays RNA-dependent ATPase activity, an activity which is stimulated by a variety of RNAs including mRNA, rRNA, tRNA, poly(A) and poly(U) (Grifo, *et al.*, 1984). Further studies revealed that this activity is specific for single-stranded RNA; double stranded RNA and single stranded DNA do not activate the ATPase activity (Abramson, *et al.*, 1987).

Perhaps the most important activity of eIF4A with respect to RNA helicases is ATP-dependent RNA unwinding. Based on a nuclease sensitivity test using purified reovirus mRNAs, Ray, *et al.* (1985) showed that in the presence of ATP, eIF4A greatly increased the nuclease sensitivity of these RNAs suggesting that the enzyme unwound duplex mRNA

structure. Later studies utilizing synthetic RNA substrates provided direct evidence of ATP-dependent unwinding activity (Rozen, *et al.*, 1990). Interestingly, Rozen, *et al.* (1990) were also able to show that eIF4A exhibits bidirectional (5' to 3' and 3' to 5') helicase activity, a unique feature among known helicases, which normally only unwind in a 3' to 5' direction (Pause and Sonenberg, 1993).

Although eIF4A is autonomously capable of performing each of these biochemical activities, it functions more efficiently as part of the eIF4F complex (Grifo, *et al.*, 1984; Ray, *et al.*, 1985). In both forms, this enzyme is stimulated by an additional factor, eIF4B (Abramson, *et al.*, 1987). eIF4B is an RNA binding protein possessing a ribosome-dependent ATPase activity (Méthot, *et al.*, 1994). While its actual function in the cell is unknown, it is absolutely required for mRNA binding to the ribosome (Méthot, *et al.*, 1994).

The role of eIF4A in the cell is postulated to be to unwind mRNA secondary structure in the 5' untranslated region in order to allow binding of the ribosome. A model has been proposed in which eIF4F binds to the 5' cap of eukaryotic mRNAs via the eIF4E subunit, and then eIF4B joins. eIF4A is released from the complex and, stimulated by eIF4B, unwinds the mRNA secondary structure to single stranded RNA (Pause and Sonenberg, 1993). In this model, eIF4E and eIF4G, the remaining members of the eIF4F complex, act to deliver eIF4A to the 5' end of the mRNA (Pause and Sonenberg, 1993). As the amount of secondary structure at the 5' end of mRNA transcripts has been shown to be inversely proportional to the translation rate (Ganzoa and Louis, 1994), eIF4A may play a critical role in their translation efficiency, in that it can unwind the secondary structure in order to allow more efficient ribosome binding and thus more efficient translation initiation.

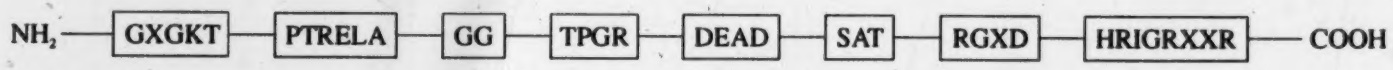
1.1.3 RNA Helicase Amino Acid Sequence Homology

As mentioned previously (Section 1.1.1), amino acid sequence comparisons are used to define RNA helicases. In the early stages of this sequence homology analysis, before specific biochemical activities were demonstrated, Linder, *et al.* (1989) identified eight proteins that showed sequence similarity to eIF4A (Section 1.1.2). The authors found several conserved amino acid motifs, including D-E-A-D (aspartic acid-glutamic-acid-alanine-aspartic acid), using single-letter amino acid codes. As a result, they defined this group of proteins as the DEAD box protein family.

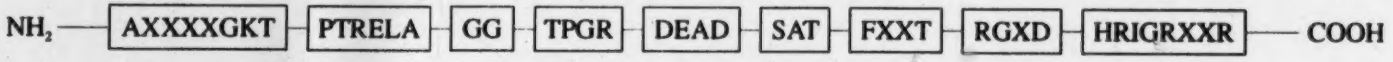
With the isolation and characterization of several other putative or established DEAD box RNA helicases, a distinct pattern of amino acid sequence has emerged. Figure 1.1 shows a schematic diagram of two typical DEAD box proteins. These proteins are characterized by a core region of 294 to 359 amino acids; it is this region that shows homology to eIF4A (Schmid and Linder, 1992). Contained within this region are a number of conserved amino acid motifs. The number of motifs ranges from eight to nine, depending on the analysis (Fuller-Pace, 1994; Pause and Sonenberg, 1993). Figure 1.1(A) shows a protein with eight motifs and 1.1(B) illustrates one with nine. The latter has an additional motif, FXXT, located between SAT and RGXD (Pause and Sonenberg, 1993).

Outside of the conserved motifs, the amino acid sequence within the core is variable. The number of amino acids between each of these motifs is relatively constant however, leading to a conservation in the spacing between each motif.

The core region is flanked by amino- and carboxy-terminal regions which vary in length (Schmid and Linder, 1992) and in amino acid sequence. It has been speculated that



A



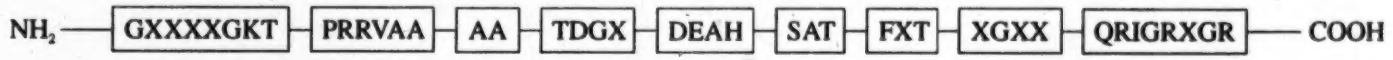
B

Figure 1.1 Schematic representations of DEAD box RNA helicases. Boxed regions indicate conserved amino acid motifs. Diagrams are not to scale. **(A)** DEAD box RNA helicase showing eight motifs (modified from Fuller-Pace, 1994). **(B)** DEAD box RNA helicase showing nine motifs (modified from Pause and Sonenberg, 1993)

these nonconserved regions could contribute to the RNA substrate specificity of each enzyme. For example, they could be responsible for sequence-specific RNA binding, accessory protein binding, or targeting signals (Schmid and Linder, 1992). In contrast, the core region is most likely responsible for the characteristic activities of RNA helicases. The conserved motifs are probably functional sites whereas the remaining core sequence provides the structural framework (Hodgman, 1988).

Although many RNA helicases contain the DEAD amino acid sequence, further study has revealed that not all members of this class of enzyme contain this motif. Two additional families, DEAH and DEXH, have been described. Figure 1.2 shows a schematic diagram of each of these families. Compared to the DEAD family, the DEAH and DEXH groups have fewer amino acid motifs and those that are present are less conserved. Similar to the DEAD family, the two latter groups of proteins are involved in processes that require the modulation of RNA secondary structure (Fuller-Pace, 1994), suggesting that the DEAD motif is not a prerequisite for RNA helicase activity.

As previously mentioned, the highly conserved amino acid motifs are most likely involved in the basic RNA helicase activities of ATP-dependent RNA binding, RNA-dependent ATPase activity and ATP-dependent RNA unwinding. In an effort to assign each of these functions to a specific motif, *in vivo* and *in vitro* mutational analyses have been performed. Schmid and Linder (1991) did an *in vivo* study of eIF4A in *Saccharomyces cerevisiae* in which the effects of mutation on growth and *in vivo* translation rates were examined. Pause and Sonenberg (1992) used purified recombinant eIF4A and eIF4B to assess mutations in each catalytic function *in vitro*.



A

B

Figure 1.2 Schematic representations of the DEAH (A) and DEXH (B) families of RNA helicases. Boxed regions indicate conserved amino acid motifs. Diagrams are not to scale (modified from Pause and Sonenberg, 1993)

The AXXGXGKT motif is a modified sequence "A" of ATP-binding proteins as described by Walker, *et al.* (1982). The presence of alanine is unusual in that most other ATP-binding proteins possess a glycine at this position (Linder, *et al.* 1989). Schmid and Linder (1991) showed that mutation of this residue to valine or aspartic acid had a negative effect on growth and *in vivo* translation rates. Substitution of a glycine residue resulted in wild-type rates (Schmid and Linder, 1991). Pause and Sonenberg (1992) concluded from their study that this motif is involved in ATP binding and that the lysine residue is particularly important in this binding. This latter conclusion is in agreement with Rozen, *et al.* (1989), who previously showed that substitution of this lysine residue eliminated ATP binding activity. These conclusions are not altogether surprising, in that the lysine residue has been shown to bind the β- and γ-phosphates of ATP (Pause and Sonenberg, 1993).

The DEAD motif is thought to represent a special form of Walker's ATPase sequence "B" (Walker, *et al.* 1982). The first aspartic acid residue binds to Mg^{2+} via a water molecule and this Mg^{2+} is complexed with the β- and γ-phosphates of ATP (Pause and Sonenberg, 1993). Schmid and Linder (1991) found that a mutation of this residue to glutamic acid resulted in non-viable cells. Pause and Sonenberg (1992) showed that a mutation in any of the four positions did not have much of a negative effect, if any, on ATP binding, but did affect ATPase and unwinding activity. They concluded that this motif is involved in ATP hydrolysis and is important in coupling ATPase and RNA helicase activities.

Schmid and Linder (1991) did not examine the SAT box, but Pause and Sonenberg (1992) determined that although ATP binding and ATPase activities were not negatively affected by mutations at both the serine and threonine residues, RNA unwinding activity

was, suggesting that this motif is essential for this function.

The only other motif that has been examined is HRIGRXXR. Schmid and Linder (1991) demonstrated that mutation in the first arginine resulted in non-viable cells except when lysine was substituted, suggesting that a positive charge is necessary at this position. Pause and Sonenberg (1992) substituted a glutamine for the histidine residue and found a decrease in ATP binding and ATPase activity as well as a loss in RNA unwinding activity. Later, Pause, *et al.* (1993) showed that a mutation in any one of the three arginines drastically reduced eIF4A binding to RNA and eliminated RNA unwinding activity. Some of these mutations also affected ATP binding and ATPase activities. It was thus concluded that this region is required for RNA binding and ATP hydrolysis (Pause, *et al.*, 1993).

At present, the functions of the remaining motifs, PTRELA, GG, TPGR, FXXT and RGXD, have not been determined, although their conservation suggests that they are somehow involved in the ATP binding, ATP hydrolysis and/or RNA unwinding activities.

An interesting point not often discussed during these mutational analyses is the role of protein conformation in RNA helicase activity. The conserved spacing between the motifs, as well as the conserved amino acids within the motifs themselves suggests that protein conformation is important. Conclusions drawn from studying the effects of single amino acid substitutions, without knowledge of concurrent effects on protein structure, must therefore be applied carefully.

1.1.4 RNA Helicase Involvement in Cellular Processes

Theoretically, RNA helicases could be involved in any process in which modulation

of RNA secondary structure is required. In fact, they have been implicated in a wide variety of cellular processes. A number of these enzymes are thought to play a role in translation initiation, much like the prototype RNA helicase, eIF4A. These include Tif1 and Tif2 from *S. cerevisiae*, functional homologues of eIF4A (Schmid and Linder, 1991), as well as SSL2 also from yeast but part of the DEAH family (Gulyas and Donahue, 1992).

Other RNA helicases are involved in ribosome biogenesis including Spb4 from *S. cerevisiae* (Sachs and Davis, 1990) and SrmB (Nishi, *et al.*, 1988) and DbpA (Böddeker, *et al.*, 1997) from *Escherichia coli*. In addition, a number of RNA helicases involved in mRNA splicing reactions have been identified in yeast. These include Prp5 (spliceosome assembly), Prp2 and Prp16 (splicing) and Prp22 (disassembly of spliceosome) (Wassarman and Steitz, 1991).

Still other RNA helicases have been implicated in development. Vasa from *Drosophila melanogaster* (Lasko and Ashburner, 1988) and An3 from *Xenopus laevis* (Gururajan, *et al.*, 1991) play a role in oogenesis and PL10 from *Mus musculus* is found only in the male germ line, suggesting a role in spermatogenesis (Leroy, *et al.*, 1989).

The above list merely gives a few examples; it is not intended to be exhaustive. In many cases, the actual function of a particular RNA helicase is unknown. This is because these members have been designated solely on the basis of amino acid sequence homology to known RNA helicases, rather than on biochemical or physiological evidence. An important point raised by this list is that RNA helicases have been found in a wide range of organisms. Examples can be found in viruses, bacteria, lower and higher eukaryotes (Fuller-Pace, 1994). In fact, RNA helicases have been found in every organism in which they have

been investigated, suggesting that these enzymes play key roles in many fundamental cellular processes. However, although a few have been shown to exhibit RNA-dependent ATPase or RNA unwinding activity, an exact RNA target has been identified for only one RNA helicase, DbpA from *E. coli* (Section 1.1.5).

1.1.5 Prokaryotic RNA Helicases

As described in the previous section, RNA helicases have been found in every organism in which they have been investigated. It is not surprising then, that a number of prokaryotic examples have been identified. Some of these examples, specifically the genes, have been discovered serendipitously as a result of sequencing large segments of genomic DNA. For example, a putative RNA helicase was found in *Bacillus subtilis* during the course of the sequencing of its genome (Genbank Accession #P42305) and another was found in *Archaea* while sequencing a fosmid clone from a DNA library prepared from an uncultivated marine Archaeon (Stein, *et al.*, 1996).

The majority of the work on prokaryotic RNA helicases however, has been done with those isolated from *E. coli*. At present, five DEAD RNA helicase genes have been identified (Kalman, *et al.*, 1991). The first was *srmB*, so named because expression of this gene at a high copy number suppresses a mutation in ribosomal protein L24 (Nishi, *et al.*, 1988). L24 is essential in the assembly of the large ribosomal subunit; the mutant protein is defective in its interaction with 23S rRNA. Nishi, *et al.* (1988) demonstrated that purified recombinant SrmB exhibits nucleic acid-dependent ATPase activity. This activity was stimulated by poly(U), poly(A), bacteriophage R17 and *E. coli* ribosomal RNA as well as DNA, especially

single-stranded. The authors also found that this enzyme is capable of ATP-independent binding of RNA.

Overproduction of SrmB may stabilize the L24 mutant protein by binding to a similar site on 23S rRNA, or it may bind to an altogether different region and protect an unstable assembly precursor from degradation (Nishi, *et al.*, 1988). Since no studies have been done using wild-type levels of SrmB, the normal physiological function of this protein remains speculative. Its ability to suppress a ribosomal protein mutation suggests that it may be involved in ribosome biogenesis.

A second *E. coli* RNA helicase was isolated based on its ability to suppress another ribosomal mutation, this time in the ribosomal protein S2, a protein that is involved in ribosome assembly (Toone, *et al.*, 1991). Toone, *et al.* (1991) named the gene *deaD*, mapped its position to 68.8 minutes on the *E. coli* map, and hypothesized that DeaD played an unknown role in translation. Three years later, Yamanaka, *et al.* (1994) isolated the gene again, this time as a multicopy supressor of a cold-sensitive growth phenotype. These authors named the gene *mssB* and suggested that it is involved in the regulation of expression of a number of different genes.

Jones, *et al.* (1996) finally assigned the gene the name *csdA* for *c*old-*s*hock *D*EAD-box protein *A*. These authors first identified this protein as a new cold-shock protein and, based on N-terminal sequencing, subsequently discovered that it was identical to DeaD. This protein was found to be a major ribosomal-associated protein at low temperature and was found to be required for optimal growth at these low temperatures (Jones, *et al.*, 1996). The authors also determined that CsdA was able to unwind a synthetic double-stranded RNA

substrate both with and without ATP, suggesting that the enzyme has helix destabilizing rather than helicase activity. They proposed that CsdA is required to unwind stable mRNA secondary structures formed at low temperature in order to increase translational efficiency (Jones, *et al.*, 1996).

A third *E. coli* RNA helicase is DbpA, found by screening an *E. coli* genomic library with a p68 cDNA probe (Iggo, *et al.*, 1990). p68 is the human SV40 large T related antigen which contains the amino acid motifs found in DEAD RNA helicases (Iggo, *et al.*, 1990). *dbpA* was mapped to 29.6 minutes on the *E. coli* chromosome and Western blotting of an *E. coli* whole cell extract with an anti-DEAD antibody revealed that this gene is expressed at very low levels (Iggo, *et al.*, 1990). Fuller-Pace, *et al.* (1993) corrected the original sequence, showing that translation starts at a GUG codon upstream from the previously predicted AUG. These authors also demonstrated that purified DbpA was able to hydrolyse ATP, but no other NTP, in the presence of total *E. coli* RNA or *E. coli* rRNA. Other RNA substrates, including total RNA from yeast and HeLa cells, *E. coli* tRNA and rabbit rRNA, did not stimulate ATPase activity. During the purification the DbpA protein however, Fuller-Pace, *et al.* (1993) showed that DbpA will bind non-specifically to any nucleic acid, suggesting that mere binding to RNA does not necessarily result in ATP hydrolysis.

The specific RNA substrate for DbpA's ATPase activity was determined to be 23S rRNA; interestingly this activity was stimulated not only by *E. coli* 23S rRNA, but by 23S rRNA from any Gram positive or Gram negative bacterial species. This was the first, and to date only identification of a specific RNA substrate for an RNA helicase. Nicol and Fuller-Pace (1995) narrowed the specificity by determining the exact site within 23S rRNA

that activates ATP hydrolysis by DbpA. The authors found that a region of 93 bases within the peptidyltransferase centre was sufficient for activation and that maintenance of secondary structure within this region was critical. The authors hypothesized, based on this result, that DbpA may play a role in translation, perhaps in the peptidyltransferase reaction itself. Alternatively, it may be involved in folding of rRNA and maintenance of the peptidyltransferase centre structure (Nicol and Fuller-Pace, 1995).

Böddeker, *et al.* (1997) later demonstrated that four additional regions of 23S rRNA were able to trigger up to 60% of DbpA's ATPase activity. Together with the 93 base region described by Nicol and Fuller-Pace (1995), these regions are spread across the 23S rRNA molecule but all are rich in secondary structure and either part of the functional core of the 50S ribosomal subunit or involved in 50S subunit assembly (Böddeker, *et al.*, 1997). As such, Böddeker, *et al.* (1997) suggested that DbpA performs a general function in the assembly process of this subunit rather than a specific role in peptidyltransferase activity.

In support of this conclusion was the observation that the RNA unwinding activity is not dependent upon ATP hydrolysis, nor does it discriminate between 23S and 16S rRNA (Böddeker, *et al.*, 1997). They also demonstrated that RNA binding is not dependent on ATP hydrolysis and that ATP binding does not require RNA binding. As a result, they suggested that DbpA has two RNA binding domains, one that is 23S rRNA-specific and coupled to ATP hydrolysis and the other that is non-specific for RNA (Böddeker, *et al.*, 1997).

A fourth *E. coli* RNA helicase is RhlB, and was found by sequencing of the 85 minute region of this organism's genome (Kalman, *et al.*, 1991). The authors could not

determine a function for this gene; it appeared to be necessary for viability only in some genetic backgrounds (the putative deletion strains constructed in that study were later found to contain a copy of *rhlB*). Study of the *E. coli* multicomponent ribonucleolytic complex, however, revealed that RhlB is a member of this complex (Miczak, *et al.*, 1996; Py, *et al.*, 1996). This complex also contains RNase E, DnaK, polynucleotide phosphorylase and enolase (Miczak, *et al.*, 1996). Also called the degradosome, it is important for RNA processing and mRNA degradation (Py, *et al.*, 1996). Py, *et al.* (1996) hypothesized that RhlB acts within this degradosome to unwind RNA secondary structures that block the processive activity of polynucleotide phosphorylase, a 5' to 3' exonuclease whose activity is impeded by RNA secondary structure. In contrast, Miczak, *et al.* (1996) speculated that RhlB unwinds RNA duplexes that inhibit the action of Rnase E. Whichever specific nuclease is involved, RhlB does appear to play a key role in RNA turnover. Interestingly, RhlB was able to bind RNA but did not show any detectable ATPase activity, suggesting that the enzyme requires other degradosome components for full activity (Py, *et al.*, 1996). The fully assembled degradosome itself did show ATPase activity *in vitro* which was stimulated by the addition of RNA (Py, *et al.*, 1996).

The fifth *E. coli* RNA helicase is RhlE, found by PCR screening of genomic DNA using oligonucleotide primers corresponding to the conserved motifs (Kalman, *et al.*, 1991). Ohmori (1994) isolated the full-length gene and determined its complete nucleotide sequence. The gene was mapped to the 17.8 minute region of the *E. coli* chromosome. In an effort to determine the function of RhlE, Ohmori (1994) created a null mutant. He found that this mutant showed no difference in growth rate when compared to wild-type and

concluded from this phenotype that RhlE is not required for growth. As a result, the function of this gene remains unknown.

As mentioned previously, a number of prokaryotic RNA helicase genes have been described. Although a putative cyanobacterial RNA helicase has been identified as a result of sequencing the *Synechocystis* sp. strain PCC 6803 genome (Kaneko, *et al.*, 1995), no further description of a cyanobacterial RNA helicase has been published.

1.2 CYANOBACTERIA

1.2.1 General Characteristics of Cyanobacteria

The cyanobacteria are a heterogeneous group of Gram negative photosynthetic prokaryotes (Stanier and Cohen-Bazire, 1977). These organisms were originally thought to be a class of eukaryotic algae, the blue-green algae, and as a result, were classified under the Botanical Code. When it was determined that they were in fact prokaryotes, efforts were made to classify them under the Bacteriological Code (Rippka, *et al.*, 1979). Currently, cyanobacteria are classified into five provisional sub-groups (Rippka, *et al.*, 1979). Section I comprises unicellular forms that reproduce by binary fission or budding. Section II also includes unicellular cyanobacteria, but only those that reproduce by multiple fission or by both multiple fission and binary fission. The filamentous cyanobacteria compose the remaining three sections. Nonheterocystous filamentous organisms are found in Section III, heterocystous filamentous forms comprise Section IV, and branching filamentous forms that produce heterocysts are found in Section V. Thus from morphology alone, it can be seen that the cyanobacteria are a very diverse group.

Cyanobacterial diversity is also evident in the wide range of areas which they inhabit. These organisms are found in marine and freshwater environments, thermal waters and terrestrial areas (Fay, 1983). These terrestrial areas can include shorelines, salt marshes and mangroves (Whitton and Potts, 1982), as well as tropical and polar soils (Fay, 1983). Acidic environments, however, appear to be devoid of cyanobacterial growth (Fay, 1983).

DNA base composition and genome size of the cyanobacteria also varies greatly. The G+C content can range from 35 to 71% (Herdman, *et al.*, 1979a), and the genome size ranges from 1.6×10^9 to 8.6×10^9 daltons (Herdman, *et al.*, 1979b). In addition, cyanobacteria are polyploid (Haselkorn, 1991). Taken together, these characteristics show that cyanobacteria are indeed an extremely diverse, heterogeneous group of prokaryotes.

All cyanobacteria are characterized by their ability to undergo oxygenic photosynthesis. Researchers have speculated that they were the first organisms on earth to evolve free molecular oxygen (O_2), a theory strengthened by the discovery of cyanobacterial fossils dating over three billion years ago (Fay, 1983). The release of free O_2 into the earth's atmosphere marked a shift from anaerobic to aerobic metabolism and the subsequent evolution of higher plants and animals (Fay, 1983). In fact, it has been determined that cyanobacteria are the ancestors of at least some of the plastids found in modern-day plants (Gray and Doolittle, 1982).

Cyanobacteria produce only the *a* form of chlorophyll, unlike higher plants which produce both the *a* and the *b* forms. The former has been found to function almost exclusively in cyclic photophosphorylation and not in the production of O_2 (Stanier and Cohen-Bazire, 1977). The major cyanobacterial light harvesting pigments are the

phycobiliproteins, of which all cyanobacteria have at least three types: phycocyanin, allophycocyanin, and allophycocyanin B (Stanier and Cohen-Bazire, 1977). Many, but not all cyanobacteria also produce phycoerythrin, which allows these organisms to undergo the process of chromatic adaptation. This is a mechanism in which pigment composition, particularly the phycoerythrin to phycocyanin ratio, is varied in response to the wavelengths of light to which the organisms are exposed (Stanier and Cohen-Bazire, 1977). Although there are a number of differences in the photosynthetic apparatus between cyanobacteria and higher green plants, these bacteria do present a simple model for the study of the photosynthetic process.

1.2.2 The Genus *Anabaena*

Anabaena is one of several genera which comprise the heterocystous filamentous group of cyanobacteria; Section IV in the Rippka, *et al.* (1979) classification scheme. All members of this genus are obligate photoautotrophs (Rippka, 1979), and their DNA base composition ranges from 38 to 44% G+C (Herdman, *et al.*, 1979a). Certain strains are able to produce akinetes, spores which are produced under adverse environmental conditions (Rippka, *et al.*, 1979). All strains are capable of heterocyst formation and aerobic nitrogen fixation (Section 1.1.4).

1.2.3 Genetic Systems in *Anabaena*

For many years, the genetic analysis of cyanobacteria in general and in filamentous forms like *Anabaena* in particular had been hampered by a lack of efficient genetic systems.

This has recently changed however, with the development of techniques for gene transfer, mutagenesis and gene reporter systems.

As a reproducible transformation system has not yet been described for the filamentous cyanobacteria, conjugation is the most widely used method for the introduction of DNA into *Anabaena*. This process involves the mobilization of a cargo plasmid, carrying the DNA of interest, by a broad host range conjugal plasmid. The conjugal plasmid used is an IncP plasmid, usually RP4 or one of its derivatives (Elhai and Wolk, 1988). The cargo plasmid can be one of two types; those that replicate in the cyanobacterium and those that do not. Wolk, *et al.* (1984) constructed a number of vectors of the former type, using pBR322 to obtain an *E. coli* replicon and the *bom* (*b*asis *o*f *m*obility) site, and pDU1 of *Nostoc* to obtain a cyanobacterial replicon. All of these vectors can be transferred from *E. coli* to *Anabaena* using the conjugal apparatus of RP4. Subsequently, a number of other cloning vectors of both types have been developed (Thiel, 1994).

Although conjugation has been shown to be an effective method to transfer DNA to cyanobacteria, its efficiency can be dramatically decreased by endogenous restriction enzyme activity. For this reason, a number of helper plasmids have been constructed which encode methylase genes for several *Anabaena* restriction activities (Elhai and Wolk, 1988; Elhai, *et al.*, 1997). When these plasmids are propagated in the same *E. coli* strain as the cargo plasmid, the result is methylation of the appropriate restriction sites in the cargo plasmid, thereby preventing the endonucleolytic attack that decreases the conjugal efficiency.

Electroporation has also been described in *Anabaena* (Thiel and Poo, 1989). The efficacy of this method, and therefore its usefulness, remain to be proven however.

Many protocols for chemical and UV-induced mutagenesis have been described (Golden, 1988); these have since been superseded by targeted mutagenesis in which a specific site can be mutated. For example, the gene of interest is disrupted *in vitro* by insertion of a selectable marker, then re-introduced into *Anabaena* in a vector that cannot replicate in the cyanobacterium. The only way the inserted marker can persist is if homologous recombination occurs between the DNA in the chromosome and that on the plasmid. One difficulty in *Anabaena* is that single crossover events occur at a much higher frequency than double crossover events (Cai and Wolk, 1990). In the former, the entire plasmid is integrated and two copies of the gene, one inactivated and one wild type, are present in the genome. In the latter, gene replacement occurs and only the inactivated copy is found in the genome. In an effort to select for double recombinants, Cai and Wolk (1990) have created a cloning vector that contains the conditionally lethal *sacB* gene from *Bacillus subtilis*. When *Anabaena* cells expressing this gene are grown in the presence of 5% sucrose on solid medium, they lyse and die. Thus, single crossover events can be screened out simply by growing exconjugants on 5% sucrose media.

Two further barriers to complete gene disruption exist in *Anabaena*. One arises from the filamentous nature of these organisms. Not all cells in a filament will be mutated and as a consequence a colony arising after plating will be a mixture of mutant and wild-type cells (Golden, 1988). Physical disruption of filaments prior to and after mutagenesis can minimize this effect (Golden, 1988). The second barrier to complete disruption is a result of their polyploidy (Haselkorn, 1991). It can be assumed that during recombination, not all chromosomes in a given cell will be mutated, therefore sufficient time must be allowed for

the segregation of mutated and non-mutated chromosomes, combined with numerous subculturings of single colonies. Complete segregation, that is, complete disruption, will not occur if the gene of interest is essential for viability (Haselkorn, 1991).

Studies in *Anabaena* genetics have advanced through the use of reporter systems to study gene expression. *lacZ* and *luxAB* have both been used with success in this organism. Elhai and Wolk (1990) utilized a *luxAB* reporter to localize nitrogenase (*nifHDK*) and photosynthetic (*rbcLS*) gene expression during heterocyst differentiation and a *lacZ* reporter to follow a time course of nitrogenase gene expression under inducing conditions. Bauer and Haselkorn (1995) created two unique reporter vectors to examine differential gene expression in heterocysts and vegetative cells of *Anabaena*. These vectors contained either a promoterless *nifHDK* operon (Section 1.2.4) followed by a promoterless *cat* gene (encoding chloramphenicol acetyltransferase) or the promoterless *cat* gene followed by the promoterless *nifHDK* operon. Harboured in a *nifH* deletion strain, the ability to grow on N_2 as the sole source of nitrogen serves as a marker for promoter activity in heterocysts while chloramphenicol resistance can be used to measure promoter activity in vegetative cells. From these examples it can be seen that a number of different genes and constructs can be used as reporters in *Anabaena*.

1.2.4 *Anabaena* sp. Strain PCC 7120

Anabaena sp. strain PCC 7120 is perhaps the most commonly used and studied strain of this genus. Through physical mapping of the genome, the chromosome size of this organism was determined to be 6.4 megabases (Mb) (Bancroft, *et al.*, 1989). These mapping

experiments also indicated that *Anabaena* sp. strain PCC 7120 DNA is resistant to cleavage by many restriction endonucleases, a phenomenon attributable in part to methylation and also to counterselection of restriction endonuclease cleavage sites (Bancroft, *et al.*, 1989).

Anabaena sp. strain PCC 7120 also appears to contain three large plasmids, designated α, β and γ of sizes 410, 190 and 110 kb, respectively (Bancroft, *et al.*, 1989). The function of these plasmids is unknown.

Further analysis of the *Anabaena* genome has revealed that a number of DNA elements are present. Six insertion sequences (IS) have been identified (Cai and Wolk, 1990; Thiel, 1994). IS*892* is 1.7 kb in size (Cai and Wolk, 1990), contains terminal repeats, and generates an 8 bp repeat in the target (Thiel, 1994). This element is likely an actively transposing IS element (Thiel, 1994). IS*895* has 30 bp terminal inverted repeats but does not duplicate the target site; this element may be more stable (Thiel, 1994). The four remaining IS elements, IS*893*, IS*894*, IS*897* and IS*898* have not yet been characterized.

The *Anabaena* genome, in addition to insertion sequences, also contains a number of repetitive sequences known as short tandemly repeated elements (STR), which have also been found in other cyanobacteria. Mulligan and Haselkorn (1989) reported seven copies of a 7 bp directly repeating sequence with a consensus sequence of CCCCAAT between the *nifB* and *fdxN* genes, and six copies between *nifS* and *nifU* (see below). These same sequences were found in at least 25 other locations in the genome (Mulligan and Haselkorn, 1989). The function of these STRs is unknown; it has been hypothesized that they may be a DNA-binding protein target for chromosome maintenance, regulate transcription termination, or act as binding sites for proteins involved in mRNA stability and processing

(Jackman and Mulligan, 1995). Most of these elements have been found outside putative coding regions. An exception was discovered in the *Anabaena nifJ* gene. Within the *nifJ* coding region, five tandemly repeated copies of the sequence CCCCGAT plus one nucleotide were discovered (Buikema and Haselkorn, 1993). The origin and function of these repeats is the subject of further investigation.

Anabaena sp. strain PCC 7120 was originally thought to encode only two restriction endonucleases, *Asp*(7120)I and *Asp*(7120)II; these enzymes are isoschizomers of *Ava*I and *Ava*II, respectively. Recently however, Elhai, *et al.* (1997) found that a third restriction activity existed, *Asp*(7120)III, an isoschizomer of *Ava*III. As a result of this discovery, the authors determined that *Anabaena* sp. strain PCC 7120 and *Anabaena* sp. strain PCC 7118 were identical; the division was originally made because no *Ava*III activity could be found in the former strain (Elhai, *et al.*, 1997).

Nitrate and ammonium can be utilized by all cyanobacteria as a source of inorganic nitrogen (Cai and Wolk, 1997b). However only certain genera, among them *Anabaena*, are capable of growth on atmospheric dinitrogen by the process of nitrogen fixation. Nitrogen fixation is extremely oxygen-sensitive, as the nitrogenase enzyme complex is oxygen-labile and inactivated irreversibly upon exposure to air (Haselkorn, 1992). A paradox arises in these organisms: how can they fix nitrogen in an environment in which O_2-evolving photosynthesis is taking place? Certain filamentous forms (Groups IV and V [Rippka, *et al.*, 1979]) circumvent the problem by producing heterocysts, terminally differentiated cells that have an anaerobic interior (Cai and Wolk, 1997a). Heterocysts are formed in the absence of a source of combined nitrogen and complete differentiation requires approximately 24 h

(Haselkorn, *et al.*, 1986). Roughly 10% of *Anabaena* vegetative cells differentiate into heterocysts (Cai and Wolk, 1997b). The environment within the heterocyst is made anaerobic by two mechanisms: the addition of a polysaccharide and glycolipid envelope, which limits O_2 diffusion into the cell, and the inactivation of Photosystem II, the O - evolving component of photosynthesis (Buikema and Haselkorn, 1993). Overall, the differentiation of a vegetative cell into a heterocyst requires the ordered expression of many genes (Lammers, *et al.*, 1986). In fact, it has been determined that over 1000 genes are differentially expressed in heterocysts (Buikema and Haselkorn, 1993).

One of the most interesting features of *Anabaena* nitrogen fixation is the rearrangement of the nitrogen fixation genes. This phenomenon was the first example of a prokaryotic developmentally regulated genome rearrangement (Haselkorn, *et al.*, 1986) and is most well defined in *Anabaena* sp. strain PCC 7120. Two such rearrangements occur, and while they both occur quantitatively in all heterocyst genomes late in heterocyst development (Golden, *et al.*, 1987), they are not obligatorily coupled (Haselkorn, 1992). Each rearrangement involves the excision of a piece of DNA that interrupts a nitrogen fixation operon and occurs through site-specific recombination using an enzyme encoded on the excised DNA (Buikema and Haselkorn, 1993). The first rearrangement to be discovered involved the *nifHDK* operon, which encodes the structural proteins of the nitrogenase enzyme complex (Buikema and Haselkorn, 1993). An 11 kb element interupts the *nifD* coding region and prevents transcription of *nifK* from the *nifH* promoter. Excision occurs through recombination between 11 bp repeats at the ends of the element and is catalyzed by XisA, encoded on the element itself (Lammers, *et al.*, 1986). This recombination results in

a 11 kb nonreplicating circular molecule and a restored *nifHDK* operon (Buikema and Haselkorn, 1993).

The second gene rearrangement occurs within the *nifB-fdxN-nifS-nifU* operon. *nifB* is involved in Fe-Mo cofactor synthesis, *fdxN* encodes a bacterial type ferredoxin, and *nifS* and *nifU* have unknown functions (Mulligan and Haselkorn, 1989). A 55 kb element interrupts the *fdxN* reading frame and is excised through site-specific recombination between 5 bp repeats by the action of XisF (Carrasco, *et al.*, 1994). In this case, a 55 kb circle is created and the operon is restored. Thus, nitrogen fixation can only occur when both elements are removed. There is evidence that the two rearrangements can be regulated independently: in *Anabaena* sp. strain PCC 7120 cultures bubbled with argon (to simulate anaerobic conditions), the 55 kb excision occurred, but the 11 kb excision did not (Golden, *et al.*, 1988). At present, the regulation of these rearrangements is still under investigation.

The function of the interrupting elements is unclear. A survey of several cyanobacterial species suggests that the 11 kb element is generally present but the 55 kb element is not (Carrasco and Golden, 1995). The latter finding suggests that the 55 kb element is dispensable. Selective pressure of an unknown nature may account for maintenance of both elements (Carrasco and Golden, 1995). Several hypotheses for the function of the 11 kb element have been put forward. These include: conferring immunity to cyanophages, encoding a restriction activity against foreign DNA and encoding a DNA damage repair enzyme (Haselkorn, 1992). Sequencing of this element has revealed no data to support any of these suggestions.

1.2.5 Use of *Anabaena* sp. Strain PCC 7120 as a Model System

Anabaena sp. strain PCC 7120 presents an attractive model system for the study of many cellular processes. This strain is well characterized on both genetic and physical levels and, as mentioned in Section 1.2.3, techniques have been developed with which to characterize and manipulate them further. For example, the conjugation system for transferring DNA from *Escherichia coli* to *Anabaena* is widely used (Elhai and Wolk, 1988), a number of shuttle vectors capable of replicating in both *E. coli* and *Anabaena* have been created (Wolk, *et al.*, 1984), and gene disruption and reporter systems have been established (Thiel, 1994).

Perhaps the most enticing reason to use this organism as a model system is because of its physiology. Although differences exist, the fundamental process of photosynthesis, studied in *Anabaena* sp. strain PCC 7120, can be applied to higher plants. Heterocyst differentiation presents an example of pattern formation in response to environmental stimuli and nitrogenase gene rearrangement can be examined to elucidate developmentally regulated genome rearrangements. Overall then, because this organism undergoes these and other well defined adaptive and developmental pathways, while at the same time being a prokaryote, cellular processes can be studied with relative ease.

1.3 THESIS OBJECTIVES

In this thesis, I report the cloning and characterization of a cyanobacterial RNA helicase gene. This project was undertaken for a number of reasons. First, as reported previously, a cyanobacterial RNA helicase has not been extensively described in the

literature. The isolation and, more importantly, the characterization of such a gene would further support the idea that these enzymes are widespread in nature and required for essential cell processes. Second, a partial clone was available with which to screen an existing *Anabaena* genomic DNA library; this clone was obtained through PCR of *Anabaena* genomic DNA using oligonucleotide primers homologous to the conserved RNA helicase motifs (W. Magee, undergraduate thesis). Finally, an RNA helicase could be isolated that is involved in a developmental or environmentally regulated pathway, like heterocyst formation or nitrogenase gene rearrangement. This finding could lead to greater insight into the pathway itself as well as show a specific role for an RNA helicase. Once an RNA helicase gene is cloned, it could be characterized in terms of its nucleotide and predicted amino acid sequence, overexpression of the protein could allow for its purification and eventual use in enzyme assays and antibody production, and *in vivo* inactivation could provide insight into the physiological importance of this gene within the cell.

CHAPTER TWO
MATERIALS AND METHODS

2.1 BACTERIAL STRAINS AND GROWTH CONDITIONS

The bacteria used in this study, their relevant genotypes, and their sources are listed in Table 2.1. Throughout the text, bacteria harboring plasmids will be written with the strain first followed by the plasmid(s) in parentheses [for example, HB101 (pRL623) denotes *Escherichia coli* strain HB101 containing plasmid pRL623]. *Anabaena variabilis* UTCC 387 was maintained on plates composed of BG-11 (Allen, 1968), containing 1% (w/v) Difco bacto agar in Coldstream incubators at 30°C under constant illumination (30 μmoles of photons/m^2/sec). Liquid cultures of *A. variabilis* UTCC 387 in BG-11 were aerated by shaking at 250 rpm and bubbling with air. BG-11 media were supplemented, where required, with antibiotics at the following concentrations: spectinomycin 2 μg/mL/ streptomycin 2 μg/mL; ampicillin 2 μg/mL.

E. coli JM109 cultures were grown in Trypticase® soy broth (TSB) (Becton Dickinson) supplemented with 2% (w/v) glucose and maintained on plates containing 1.2% (w/v) Difco agar. Where required, ampicillin was added to 100 μg/mL.

All other *E. coli* strains were grown in LB media (5 g/L yeast extract, 10 g/L bacto tryptone, 5 g/L NaCl) and maintained on plates containing 1.2% (w/v) Difco agar. LB media were supplemented with antibiotics, where required, at the following concentrations: ampicillin 100 μg/mL; chloramphenicol 25 μg/mL; ampicillin 15 μg/mL/spectinomycin 15 μg/mL; chloramphenicol 12.5 μg/mL/spectinomycin 15 μg/mL.

All *E. coli* strains were grown at 37°C.

TABLE 2.1: Bacterial Strains

Strain	Relevant Genotype	Reference/Source	Use
Anabaena variabilis UTCC 387 (equivalent to *Anabaena* sp. strain PCC[a] 7120)	Wild type	University of Toronto Culture Collection (UTCC)	Study subject
Escherichia coli DH5α	*supE44 ΔlacU169 (ø80 lacZΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1*	Sambrook, *et al.*, 1989	Cloning
Escherichia coli JM109	*recA1 supE44 endA1 hsdR17 gyrA96 relA1 thi Δ(lac-proAB) F' [traD36 proAB$^+$ lacI^q lacZΔM15]*	Sambrook, *et al.*, 1989	Protein overexpression
Escherichia coli HB101	*supE44 hsdS20($r_B^-m_B^-$) recA13 ara-14 proA2 lacY1 galK2 rpsL20 xyl-5 mtl-1*	Sambrook, *et al.*, 1989	Triparental mating/Gene inactivation
Escherichia coli JE2571	F$^-$ *pil fla Str Nal Rif*	Frost and Paranchych, 1988	Triparental mating/Gene inactivation

[a]Pasteur Culture Collection

TABLE 2.2: Parent Plasmids

Plasmid	Reference/Source	Selective Markers	Use
pBluescript KS+	Stratagene	Ampicillin	Cloning
pGEX-2T	Pharmacia	Ampicillin	Protein overexpression in *E. coli*
pBR322	Bolivar, *et al.*, 1977	Ampicillin/Tetracycline	Cloning/Triparental mating/Gene inactivation
RP4	Datta, *et al.*, 1971	Ampicillin/Tetracycline/Kanamycin	Conjugative plasmid/Triparental mating/Gene inactivation
pRL623	Elhai, *et al.*, 1997	Chloramphenicol	Helper plasmid/Triparental mating/Gene inactivation

TABLE 2.3: Plasmid Constructions

Plasmid	Parent Plasmid	Selective Markers	Insert	Cloning Strategy	Use
pWM75	pBluescript KS+	Amp	4 kb	*Anabaena Xba*I genomic clone bank	Original clone
pWM752	pBluescript KS+	Amp	936 bp	*Eco*RV/*Eco*RV from pWM75	Sequencing
pWM753	pBluescript KS+	Amp	2418 bp	*Hinc*II/*Hinc*II from pWM75	Construction of pBRINAC
pWM754	pBluescript KS+	Amp	1028 bp	*Eco*RV/*Hinc*II from pWM75	Sequencing
pWM755	pBluescript KS+	Amp	1487 bp	*Acc*I/*Hinc*II from pWM75	Sequencing
pGX29	pGEX-2T	Amp	1.4 kb	Site-directed mutagenesis to create *Bam*HI/*Eco*RI ends of *crhA* ORF	Protein overexpression
pKSINAC	pBluescript KS+	Amp/Spc/Sm	4418 bp	pWM753 *Acc*I digest and fill in + Spc^R/Sm^R cassette	Construction of pBRINAC
pBRINAC	pBR322	Amp/Spc/Sm	4418 bp	*Hinc*II/*Hinc*II from pKSINAC	Triparental mating/ Gene inactivation

Abbreviations: Amp, Ampicillin; Spc, Spectinomycin; Sm, Streptomycin

2.2 PURIFICATION OF DNA

2.2.1 Small Scale Purification of *E. coli* Plasmid DNA

Small scale purification of plasmid DNA was performed using the TENS method (Zhou, *et al.*, 1990), with minor modifications. 1.5 mL of an overnight culture were transferred to an eppendorf tube and the tube spun in a microcentrifuge for 10 sec to pellet the cells. The supernatant was decanted, leaving 50-100 μL and the cell pellet resuspended by vortexing. 300 µLof TENS (0.1 N NaOH; 0.5% [w/v] SDS; 10 mM Tris, pH 8.0; 1 mM EDTA, pH 8.0) was added followed by 150 μL of 3 M sodium acetate, pH 5.2, and the mixture vortexed for 2 sec after each addition. Cellular debris was pelleted by microcentrifugation for 5 min and the supernatant transferred to a clean tube. After addition of 900 μL of ice cold 100% ethanol, the solution was mixed and immediately spun at room temperature for 2 min. The DNA was washed twice with ice cold 70% (v/v) ethanol, allowed to air dry, and resuspended in 150 μL sterile distilled H_2O.

2.2.2 Large Scale Purification of *E. coli* Plasmid DNA

Plasmid DNA was purified from 500 mL *E. coli* cultures using the QIAGEN Plasmid Maxi kit according to the manufacturer's protocol with one modification. The cellular debris obtained after lysis was pelleted by centrifugation at 6000 rpm for 1 h at 4°C in a Beckman JA-14 rotor. The DNA was resuspended in sterile distilled H_2O at a final concentration of 1 µg/µL.

2.2.3 Purification of *Anabaena variabilis* UTCC 387 Chromosomal DNA

For large-scale purification of chromosomal DNA from wild type *A. variabilis* UTCC 387, cells from a 250 mL log phase culture were harvested at 6700 rpm for 15 min at room temperature in a Beckman JA-14 rotor. The supernatant was decanted and the cells washed once in 8 mL of SE buffer (120 mM NaCl; 50 mM EDTA, pH 8.0). The cell pellet was resuspended in 8 mL of SE buffer and the suspension transferred to a 50 mL screw-cap tube. Lysozyme powder (Boehringer Mannheim) was added to 3 mg/mL and the mixture incubated at 37°C for 70 min with occasional shaking. Once cell lysis was complete, as evidenced by a very viscous culture, SDS was added to 0.1% (w/v). The solution was mixed gently and incubated at 37°C for 45 min, followed by a 5 min incubation at 60°C. $NaClO_4$ was added to 0.4 M and the solution mixed well. The solution was extracted twice with phenol and twice with chloroform:isoamyl alcohol (24:1), alternating between screw cap tubes for shaking and corex tubes for centrifugation. Centrifugation was performed at 4°C in a Beckman JA-20 rotor at 5000 rpm for 30 min (phenol extractions) or 5 min (chloroform:isoamyl alcohol extractions). Two volumes of 100% ethanol were added to the final aqueous layer and the DNA collected by centrifugation at 8500 rpm for 25 min at 4°C. The pellet was air dried, resuspended in 1 mL sterile distilled H_2O, and transferred to an eppendorf tube. The solution was incubated at 37°C for 1 h with 50 μg/mL RNase. Following addition of NaCl (0.2 M), the solution was extracted once each with phenol, phenol:chloroform (1:1), and chloroform:isoamyl alcohol (24:1). DNA was precipitated by the addition of 2.5 volumes of 100% ethanol and collected by microcentrifugation for 20 min at 4°C. The pellet was washed twice in 70% (v/v) ethanol, allowed to air dry and gently resuspended in 300 μL sterile distilled H_2O.

For small-scale purification of chromosomal DNA from inactivation mutants of *A. variabilis* UTCC 387, a modification of the procedure by Cai and Wolk (1990) was used. Cells from a 25 mL log phase culture were harvested at 8500 rpm for 10 min in a Beckman JA-20 rotor at room temperature. The supernatant was decanted and the pellet resuspended in 400 μL 10 mM Tris; 0.1 mM EDTA, pH 7.5. An aliquot of this cell suspension (500 μL) was added to an eppendorf tube containing 150 μL sterile glass beads (0.5 millimeter diameter, Biospec Products). 20 μL of 10% (w/v) SDS and 450 μL phenol:chloroform (1:1) were added and the mixture subjected to 4 cycles of vigorous vortexing for 1 min followed by chilling on ice for 1 min. The mixture was then spun in a microcentrifuge for 10 min at 4°C. The supernatant was removed to a clean eppendorf tube and extracted once each with phenol:chloroform (1:1) and chloroform:isoamyl alcohol (24:1). DNA was precipitated for 1 h at -20°C after the addition of one-tenth volume of 3 M sodium acetate, pH 5.2, and 2.5 volumes of 100% ethanol. DNA was collected in a microcentrifuge at 4°C for 30 min, washed twice with 70% (v/v) ethanol, air dried and resuspended in 120 μL sterile distilled H_2O.

DNA concentrations were determined by measuring the absorbance at 260 nm (A_{260}), where one A_{260} unit is equivalent to 50 μg/mL DNA.

2.2.4 Purification of *E. coli* Chromosomal DNA

Chromosomal DNA was isolated from a 3 mL overnight culture using a modification of Marmur's procedure (1961). Cells were harvested in a microcentrifuge and resuspended in 200 μL saline-EDTA (0.15 M NaCl; 0.1 M EDTA). SDS was added to 3% (w/v) and the

mixture incubated at 60°C for 10 min. $NaClO_4$ was added to 1 M and the volume brought to 750 μL by the addition of saline-EDTA. The mixture was extracted twice with phenol and twice with chloroform:isoamyl alcohol (24:1). Two volumes of 100% ethanol were added and the DNA spooled on a heat-sealed Pasteur pipette. The DNA was washed twice in 70% (v/v) ethanol, allowed to air dry and resuspended in 375 μL sterile distilled H_2O. RNase (50 μg/mL) was added and the solution incubated at 37°C for 30 min. NaCl was added to 0.2 M and the solution extracted once each with phenol, phenol:chloroform (1:1), and chloroform:isoamyl alcohol (24:1). Two volumes of 100% ethanol were added to the final aqueous layer and the DNA precipitated at -20°C for 1 h. DNA was collected in a microcentrifuge at 4 °C for 30 min, washed twice with 70% (v/v) ethanol, allowed to air dry and resuspended in a final volume of 200 μL sterile distilled H_2O. The DNA concentration was determined spectrophotometrically by measuring its absorbance at 260 nm.

2.3 MANIPULATION OF DNA

2.3.1 Digestion and Gel Electrophoresis of DNA

DNA was digested using restriction endonucleases obtained from Boehringer Mannheim, New England Biolabs and Promega, according to the manufacturers' recommendations.

DNA fragments were separated on agarose gels (0.7 - 1.2% [w/v]) using either 0.5X TBE (45 mM Tris-borate; 1 mM EDTA) or 1X TAE (40 mM Tris-acetate; 1mM EDTA) as the buffer system. One-fifth volume of 5X DNA loading buffer (30% [w/v] sucrose; 0.125% [w/v] bromphenol blue; 5 mM EDTA, pH 8.0) was added to each sample prior to loading;

RNase (20 μg/mL) was also included when loading small-scale DNA samples. Gels were electrophoresed at constant voltage for various lengths of time, stained in an ethidium bromide solution (10 μg/mL), and the DNA fragments visualized on a UV transilluminator. Fragment sizes were determined by comparing migration distances to a standard curve constructed using lambda DNA digested with *Hin*dIII and/or pBR322 DNA digested with *Bst*NI, run simultaneously on each gel, as standards.

2.3.2 Purification of DNA from Agarose Gels

DNA fragments were purified from 1X TAE gels using the Geneclean® II method (Bio 101) according to the manufacturer's instructions, with some modifications (Smith, *et al.*, 1995). The DNA band of interest was excised from the gel, placed into an eppendorf tube and its weight determined (1 g = 1 mL). After the addition of three volumes of 6 M NaI, the tube was incubated at 55°C for 5 min, mixing every 1-2 min. 5 μL Glassmilk were added and the solution incubated at 55°C for 5 min, mixing every 1-2 min. The Glassmilk was pelleted for 5 sec in a microcentrifuge and the supernatant removed completely. The pellet was washed three times with 200 μL of New Wash (50% [v/v] ethanol; 50% [v/v] buffer [20 mM Tris-HCl, pH 7.2; 0.2 M NaCl, 2 mM EDTA]). For DNA fragments smaller than 1.5 kb, the pellet was washed by resuspension in New Wash and microcentrifugation for 5 sec. For DNA fragments larger than 1.5 kb, the pellet was washed by diffusion for 5 min followed by microcentrifugation for 5 sec. In both cases, all traces of New Wash were removed after the third wash. DNA was eluted from the Glassmilk by resuspending the pellet in 20 μL of sterile distilled H_2O and incubating at 55°C for 2-3 min. Glassmilk was

pelleted for 30 sec and the supernatant collected. The elution step was repeated once and the eluates pooled.

2.3.3 Southern Transfer and Hybridization

DNA fragments separated on agarose gels were transferred by capillary action to Hybond™ -N 0.45 micron nylon membranes (Amersham Life Science) essentially as described by Sambrook, *et al.*, 1989. Gels were soaked in denaturing buffer (1.5 M NaCl, 0.5 N NaOH) for 15-30 min with gentle shaking, rinsed with distilled H_2O and treated with neutralization buffer (0.5 M Tris, 1.5 M NaCl, pH 7.4) for 1 h. The capillary transfer was set up as follows: A piece of Whatman 3 MM paper soaked in 10X SSC (1.5 M NaCl; 150 mM sodium citrate, pH 7.0) was placed on a glass plate which sat on a pyrex baking dish filled half-way with the same buffer. The ends of the paper rested in the buffer. The gel was arranged on top of the Whatman paper, followed by the nylon membrane and two pieces of Whatman 3 MM paper, pre-soaked in 10X SSC. Paper towels were next added to a height of 10 cm followed by a glass plate and a small weight. The transfer was allowed to proceed overnight. After disassembly, DNA was fixed to the membrane by exposure to UV light using the Bio-Rad GS Gene Linker™ and the manufacturer's recommended setting.

Membranes were prehybridized at 65°C for at least 1 h in sealed plastic bags containing 1 mL prewarmed hybridization buffer (6X SSC [0.9 M NaCl; 90 mM sodium citrate, pH 7.0]; 0.5% [w/v] SDS; 5X Denhardt's solution [0.1% [w/v] Ficoll; 0.1% [w/v] PVP; 0.1% [v/v] BSA]; 0.01 M EDTA, pH 8.0) per 10 cm^2 of membrane area and 100 μg/mL boiled sonicated herring sperm DNA (Sigma). Hybridization was carried out at the

prehybridization temperature for 17-23 h in an equal volume of fresh hybridization buffer to which ^{32}P-labelled probe (10^6 cpm per mL hybridization buffer) was added. Plasmid DNA fragments, separated and purified as described above (Section 2.3.2), were utilized as probe templates. The DNA probe templates were labelled with [α^{32}P]-dCTP (Amersham Life Science) using the random primer method (Feinberg and Vogelstein, 1983). 25-50 μg DNA in a final volume of 9 μL were boiled for 5 min and immediately placed on ice for 5 min. The remaining reaction components were added in the following order: 2 μL 10X Hexanucleotide mix (Boehringer Mannheim), 3 μL dNTP mix (dATP, dGTP and dTTP, each at 0.5 mM), 5 μL [α^{32}P]-dCTP, 2 U Klenow fragment of DNA polymerase (Boehringer Mannheim). The reaction was incubated at room temperature overnight and 6 μL Blue Dextran in 0.5 M EDTA, pH 8.0 were added. Unincorporated label was then separated from labelled probe by passage over a Sephadex G-50 column and the amount of radioactivity present determined by Cerenkov counting. Immediately prior to being added to the hybridization buffer, probes were boiled for 5 min then incubated on ice for 5 min. Following hybridization, membranes were washed once in 2X SSC; 0.1% (w/v) SDS for 10 min at room temperature, followed by two to three washes in 0.2X SSC; 0.1% (w/v) SDS for 10-15 min at room temperature and, in certain cases, one wash in 0.2X SSC; 0.1% (w/v) SDS for 1 h at 55°C. Washed membranes were wrapped in saran wrap and exposed to Kodak X-OMAT™ AR X-ray film at -80°C for 2 min to 25 h. Film was developed using a Fuji RG II X-ray Film Processor.

2.3.4 *In situ* Hybridization of Bacterial Clones

For screening of *A. variabilis* UTCC 387 genomic clone banks in *E. coli* DH5α, cells were plated to give approximately 100 colonies per plate, incubated overnight at 37°C then cooled at 4°C for 30-60 min. The colonies were transferred to 0.45 micron nitrocellulose filters (Millipore) by overlaying the filters onto the agar surface for 1 min. In order to orient the filter with the plate, a 18 gauge needle was used to punch holes in a random asymmetric pattern on the edges of the filter. Filters were placed colony side up on a piece of Whatman 3 MM paper saturated with 10% (w/v) SDS for 3 min. Filters were then placed consecutively onto pieces of Whatman 3 MM paper soaked with denaturing buffer (1.5 M NaCl; 0.5 N NaOH), neutralizing buffer (0.5 M Tris, pH 7.4; 1.5 M NaCl) and 2X SSC (0.3 M NaCl; 30 mM sodium citrate, pH 7.0) for 5 min each. Filters were dried for at least 30 min on Whatman 3 MM paper and baked at 80°C under vacuum for 1-2 h. Baked filters were wet in 2X SSC for 5 min and incubated in Prewash solution (5X SSC; 0.5% [w/v] SDS; 1 mM EDTA) for 30 min at 50°C with gentle shaking. Bacterial debris was removed by rubbing the filters with a kleenex then rinsing in 2X SSC. Filters were prehybridized at 68°C for 1-2 h in a plastic dish containing 1 mL hybridization buffer (6X SSC; 0.25% [w/v] skim milk [Difco]; 0.001% [w/v] sodium azide) per 10 cm^2 of filter area. Hybridization was carried out at the prehybridization temperature for 48 h in an equal volume of hybridization buffer to which [α^{32}P]-dCTP labelled probe (10^6 cpm per mL hybridization buffer) was added. Probes were labelled as described above (Section 2.3.3). Filters were washed once in 2X SSC; 0.1% (w/v) SDS for 10 min at room temperature and once in 0.2X SSC; 0.1% (w/v) SDS for 10 min at room temperature. Washed filters were wrapped in saran wrap and exposed to a storage phosphor screen (Molecular Dynamics) for 22.5 h. Images were

retrieved from the screen using a Molecular Dynamics Phosphor Imager 445 SI and ImageQuant™ version 4.1 image analysis software (Molecular Dynamics).

2.3.5 DNA Sequencing

DNA was sequenced by the dideoxy chain termination method using the Sequenase™ version 2.0 DNA sequencing kit (United States Biochemical). Double stranded plasmid DNA, purified by either the small or large scale protocols (Sections 2.2.1 and 2.2.2), was denatured prior to sequencing. For each 20 μL reaction, 3-6 μg DNA in 18 μL sterile distilled H_2O were mixed with 2 μL fresh denaturing buffer (2 N NaOH; 2 mM EDTA) and incubated at 50°C for 5 min. DNA was precipitated at -80°C for 5 min after the addition of 14 μL 5 M ammonium acetate and 100 μL 100% ethanol, collected by microcentrifugation for 20 min at 4°C, and washed two times in 70% (v/v) ethanol. The final pellet was allowed to air dry and resuspended in 7 μL sterile distilled H_2O. Oligonucleotide primers (0.1 μg per reaction) were annealed to denatured template DNA for 12 min at 37°C in a 10 μL reaction volume containing 2 μL Sequenase™ buffer. Sequencing primers used in this study are listed in Table 2.4. After annealing, template DNA was cooled to room temperature for at least 30 min then placed on ice. Template DNA was labelled at room temperature for 5 min in a reaction containing 1 μL 0.1 M DTT, 2 μL diluted labelling mix (1:5 in distilled H_2O), 0.5 μL [α^{35}S]-dATP (Amersham Life Science), and 2 μL diluted Sequenase™ polymerase (1:13 in Sequenase™ dilution buffer). Termination reactions were carried out by adding 3.5 μL of labelled template DNA to 2.5 μL of each termination mixture (ddATP, ddCTP, ddGTP, ddTTP), prewarmed to 37°C. Reactions were incubated at 37° C for 5 min and

TABLE 2.4: Oligonucleotide Primers

Name	Sequence (5'→3')	Origin of Sequence	Use
Forward (M13 -20 primer)	GTAAAACGACGGCCAGT	pBluescript KS+	Sequencing
Reverse	AACAGCTATGACCATG	pBluescript KS+	Sequencing
GWO 22	GACTATCTTGGG	*crhA* bp 201 to 190 (antisense strand)	Sequencing
WCM 2	ACAGAGTACAGGTT	*crhA* bp -143 to -130 (antisense strand)	Sequencing
WCM 3	TTGGATTATATGT	*crhA* bp -156 to -144 (sense strand)	Sequencing
GWO 37	CCCCCTCCTACATTA	*crhA* bp 171 to 185 (sense strand)	Sequencing
GWO 35	ACCTTGAGCAGCAGA	*crhA* bp 1227 to 1213 (antisense strand)	Sequencing
WCM 1	TGTCAGTTGCTACT	*crhA* bp 913 to 900 (antisense strand)	Sequencing
GWO 36	CCATGAGCGCATTCA	*crhA* bp 301 to 315 (sense strand)	Sequencing
GWO 27	GATATTGAAAAACTG	*crhA* bp 1078 to 1092 (sense strand)	Sequencing
GWO 33	GGGCTGGCAAGCCACGTTTGGTG	pGEX-2T	Sequencing
GWO 34	CCGGGAGCTGCATGTGTCAGAGG	pGEX-2T	Sequencing
WCM 4	TCGGAGAATTCTAAAGCGAT	*crhA* bp 1408 to 1389 (antisense strand)	Mutagenesis
WCM 5	AGACATGGATCCCCAGA	*crhA* bp 6 to -11 (antisense strand)	Mutagenesis

stopped by the addition of 4 μL stop solution. For reading sequences close to the primer, labelling mix was diluted 1:10 in distilled H_2O. For reading sequences farther from the primer, 1 μL of each termination mixture and 1.5 μL Sequence Extending Mix were used in the termination reactions. Otherwise, these reactions were carried out as above. Sequencing reactions were used immediately or stored at -20°C.

Six percent (w/v) polyacrylamide gels were cast in 20x40 cm Bio-Rad Sequi-Gen® II nucleic acid sequencing cells. A plug consisting of 10 mL 6% acrylamide stock (460 g/L ultrapure urea; 1X TBE [90 mM Tris-borate; 2 mM EDTA]; 150 mL/L 40% acrylamide:bisacrylamide [Bio-Rad 19:1]); 50 μL 25% (w/v) APS and 50 μL TEMED was poured first and allowed to polymerize. Each gel was prepared by combining 50 mL 6% acrylamide stock, 75 μL 25% (w/v) APS and 75 μL TEMED and degassing the mixture by vacuum filtration before adding the TEMED. Gels were allowed to polymerize at least 1 h prior to loading. Three and one-half microlitre aliquots of the sequencing reactions, heated at 75°C for 2 min, were loaded using sharkstooth combs and the gels run at a constant power of 45 W in 1X TBE buffer (prewarmed to 77°C) for 1.5 - 10 h. Gels were fixed in 5% (v/v) methanol; 5% (v/v) acetic acid for 30 min, transferred to Whatman 3MM paper and dried under vacuum at 80°C for 1 h. To visualize DNA bands, dried gels were exposed to Kodak X-OMAT™ AR X-ray film at room temperature for 16-48 h. X-ray film was developed using a Fuji RG II X-ray Film Processor.

DNA sequence analysis was performed using the University of Wisconsin Genetics Computer Group Sequence Analysis Software package version 8.1. DNA sequences were translated into amino acid sequences using the MAP and TRANSLATE programs. The

MAP program was also used to determine restriction endonuclease cleavage sites. Amino acid sequences were compared using the BLAST and GAP programs and aligned using the PILEUP and PRETTY programs, and amino acid composition analysis was done with the PEPTIDESORT program.

2.3.6 DNA Ligation

DNA fragments, purified from agarose gels as described above (Section 2.3.2), were ligated to restriction endonuclease-digested plasmid cloning vectors at 15°C for 16-24 h using T4 DNA ligase and the manufacturer's supplied buffer (Boehringer Mannheim). For cohesive end ligations, a 3:1 insert to vector molar ratio was used and the final ligase concentration was 0.05 U/μL. For blunt end ligations, the insert to vector molar ratio was 5:1 and the final ligase concentration was 0.1 U/μL.

2.3.7 Bacterial Transformation

Competent *E. coli* DH5α and HB101 for transformation were prepared by treatment with rubidium chloride/calcium chloride solutions. An overnight culture, grown in LB broth, was diluted 1:100 into fresh LB broth and incubated at 37°C on a rotary shaker until an OD_{600} of 0.6 was reached. The culture was cooled at 4°C and the cells harvested at 4000 rpm for 10 min in a Beckman JA-14 rotor at 4°C. The cell pellet was resuspended in one-half volume ice cold 10 mM RbCl; 10 mM MOPS, pH 7.0 and incubated on ice for 10 min. The cells were centrifuged, the supernatant decanted and the pellet resuspended in one-half volume ice cold 10 mM RbCl; 0.1 M MOPS, pH 6.5; 50 mM $CaCl_2$. After incubation on ice

for 20-30 min, the cells were repelleted, resuspended in one-tenth volume RbCl; MOPS; $CaCl_2$ solution and DMSO added to a final concentration of 7% (v/v). Cells were dispensed into 1 mL aliquots and stored at -80°C. Prior to use, cells were thawed on ice for at least 30 min. For each transformation, 200 μL of competent cells were added to a prechilled eppendorf tube and 1 μg DNA added. The mixture was gently mixed, incubated on ice for 30 min and heat shocked at 42°C for 2 min. One millilitre of LB broth was added and the cells incubated for 1 h at 37°C on a rotary shaker. Aliquots (50-150 μL) were plated onto LB plates containing the appropriate antibiotics and the plates incubated at 37°C overnight. For detection of β-galactosidase activity for blue/white color selection of pBluescript KS+ (Stratagene) clones, 50 μL of a 5:1 X-gal:IPTG (2% [w/v] X-gal; 100mM IPTG) mixture was spread on plates prior to plating cells.

2.3.8 Bacterial Electroporation

Electrocompetent *E. coli* JM109 cells were prepared by diluting an overnight culture, grown in TSB from a single colony, 1:100 into 5 mL fresh TSB broth and incubating at 37°C on a rotary shaker until an OD_{600} of 0.3-0.4 was obtained. Cells were pelleted at 8000 rpm for 2 min, washed twice with 1 mL sterile dH_2O and twice with 1 mL ice-cold 15% (v/v) glycerol, and resuspended in ice-cold 15% glycerol (250 μL). For each electroporation, 40 μL of chilled cells were mixed with 1 μg DNA and added to a pre-chilled cuvette (BTX Electroporation Cuvettes Plus™). Electroporation was carried out in a Bio-Rad Gene Pulser® II using 200 Ω resistance, 2500 V voltage and a capacitance of 25 μF. One millilitre of TSB was immediately added and the cells incubated at 37°C for 1 h without shaking.

Aliquots of 100 μL were plated onto TBS plates containing the appropriate antibiotics and the plates incubated at 37°C overnight.

2.4 OVEREXPRESSION OF PROTEIN IN *E. coli*

2.4.1 Cloning of the *crhA* Open Reading Frame into Expression Vector pGEX-2T

A modification of the site-directed mutagenesis protocol by Zoller and Smith (1987) was used to create a *Bam*HI site immediately preceeding the initiator ATG and an *Eco*RI site 100 bp downstream of the stop codon of the *crhA* open reading frame in order to facilitate cloning. Serial dilutions (1:10, 1:100, 1:500, 1:1000) of a pWM75 (Table 2.3) large-scale preparation were denatured in a 20 μL reaction volume containing 2 μL denaturing buffer (2 N NaOH; 2 mM EDTA) at 50°C for 5 min. The DNA was precipitated with 14 μL 5 M ammonium acetate and two volumes of 100% ethanol for 5 min at -80°C, collected by microcentrifugation, washed twice with 70% (v/v) ethanol, air dried and resuspended in 7 μL distilled H_2O. Resuspended DNA was stored on ice until required. Oligonucleotide primers were prepared as follows: each mutagenic primer (WCM 4 and WCM 5 [Table 2.4]) was phosphorylated in a 20 μL reaction volume containing 200 pmol primer, 2 μL 10X kinase buffer, 2 μL 10 mM ATP and 2 μL diluted polynucleotide kinase (1:10 in distilled H_2O) (Boehringer Mannheim). After reactions were incubated at 37°C for 30 min, an additional 2 μL diluted polynucleotide kinase were added and the reaction incubated another 30 min. The enzyme was heat inactivated at 65°C for 10 min. The third primer, reverse sequencing primer (Table 2.4), was diluted 1:10 in distilled H_2O. To 1 μL of each chilled, denatured template DNA was added: 3 μL each phosphorylated mutagenic primer, 1 μL

diluted reverse primer, 1 μL Solution A (0.2 M Tris, pH 7.5; 0.1 M $MgCl_2$; 0.5 M NaCl; 0.01 M DTT) and 4 μL H_2O. The solutions were mixed and incubated at 37°C for 12 min. Reactions were cooled to room temperature for at least 10 min and placed on ice. 10 μL Solution C (1 μL Solution B [0.2 M Tris, pH 7.5; 0.1 M $MgCl_2$; 0.1 M DTT]; 4 μL 2.5 mM dNTP mix; 1 μL 10 mM ATP; 3U T4 ligase; 2.5 U Klenow fragment of DNA polymerase) were then added and the reactions incubated at 30°C for 1 h. An additional 10 μL Solution C were added and incubation continued at 15°C overnight. Each of the four reactions was transformed into *E. coli* DH5α as described above (Section 2.3.7), except that after incubating at 37°C for 1 h, the entire transformation mixture was inoculated into 5 mL LB broth containing 100 μg/mL ampicillin and grown overnight at 37°C on a rotary shaker.

In order to isolate *crhA* fragments flanked by *Bam*HI/*Eco*RI ends, a small-scale plasmid preparation was done on each 5 mL LB broth culture, digested with both enzymes and run on an agarose gel. A 1.4 kb band, corresponding to the fragment of interest, was gel purified as described above (Section 2.3.2) and ligated into *Bam*HI/*Eco*RI digested pGEX-2T (Pharmacia). This plasmid, designated pGX29 (Table 2.3), contains an in-frame fusion with the coding sequence for glutathione-S-transferase (GST) already located in pGEX-2T. pGX29 also encodes a thrombin cleavage site at the point of the GST/*crhA* fusion. This plasmid was confirmed by restriction analysis and partial DNA sequencing.

Purified pGX29 was electroporated into *E. coli* JM109, resulting in the strain JM109 (pGX29).

2.4.2 Batch Purification of Fusion Protein

JM109 (pGX29) was induced to overproduce the fusion protein as follows. A single colony was innoculated into 5 mL TSB containing 2% glucose and 100 μg/mL ampicillin, grown overnight and diluted 1:100 into 20 mL fresh media. The diluted culture was grown to an OD_{600} of 0.6, induced with IPTG to a final concentration of 0.5 mM and the cells grown for a further 4 h. Cells were harvested, washed in STE (10 mM Tris, pH 8.0; 150 mM NaCl; 1 mM EDTA) and stored as a pellet at -20°C.

For experiments involving the purification of soluble protein, the frozen cell pellet was thawed on ice and resuspended in 10 mL STE. Lysozyme powder (Boehringer Mannheim) was added to 100 μg/mL, the solution incubated on ice for 15 min then DTT added to 5 mM. For some purification tests, one of the following detergents was also added: 0.25-2% (w/v) sarkosyl (Sigma), 0.03% (w/v) SDS, 1% CTAB (Sigma), or 2% (v/v) TWEEN® 20 (Caledon Laboratories Ltd.). The solution was then passed through a French Pressure Cell (American Instrument Company) operating at approximately 1300 pounds per square inch. Lysed cells were centrifuged at 8500 rpm for 10 min at 4°C in a Beckman JA-20 rotor to remove debris. The supernatant was transferred to a clean 15 mL screw-cap tube and, to some samples, 2-6% TritonX-100 (Sigma) was added. Some of these samples were then incubated at 4°C for 30 min with gentle shaking. Certain samples, with or without TritonX-100 were also dialyzed against two changes of STE at 4°C using either Spectra/Por ® Molecularporous Dialysis Membranes (MWCO=6-8000) or Pierce Slide-A-Lyzer™ Dialysis Cassettes (MWCO=10,000).

Binding of the fusion protein to Glutathione Sepharose 4B (Pharmacia) was performed essentially as described in the Pharmacia GST Gene Fusion System Manual

(1993). 200 μL of a 50% Glutathione Sepharose 4B slurry in 1X phosphate-buffered saline (PBS) were added to each 10 mL cleared cell lysate. The mixture was incubated with gentle shaking at room temperature for 0.5-2 h or at 4°C for 0.5 h to overnight. Glutathione Sepharose 4B beads were collected by centrifugation at 3000 rpm in a clinical centrifuge and washed three times by the addition of 1 mL 1X PBS per 10 mL cleared cell lysate. Any bound fusion protein was eluted from the beads by adding 100 μL Glutathione elution buffer (10 mM reduced glutathione in 50 mM Tris-HCl, pH 8.0) per 10 mL cleared cell lysate, incubating at room temperature for 10 min with gentle shaking, sedimenting the beads at 3000 rpm in a clinical centrifuge, and collecting the supernatant.

For experiments aimed at purifying insoluble fusion protein, each frozen cell pellet was thawed on ice and resuspended in 10 mL STE. Lysozyme powder was added to 100 μg/mL, the mixture incubated on ice for 15 min and 5 mM DTT added. The mixture was passed one to three times through a French Pressure cell as above. The cell lysate was centrifuged at 5000 rpm for 10 min at 4°C in a Beckman JA-20 rotor. The supernatant was removed and the pellet washed in 2 mL STE or 0.25% (w/v) sarkosyl per 10 mL cell lysate. The fusion protein was extracted from the pellet by adding 2 mL extraction buffer (1.5% [w/v] sarkosyl in STE) per 10 mL cell lysate and vortexing vigorously for 5 min. The mixture was centrifuged at 5000 rpm for 10 min at 4°C in a Beckman JA-20 rotor and the supernatant collected.

For thrombin cleavage experiments, 100 μL thrombin solution (prepared by diluting thrombin [Sigma] to 0.05 cleavage units/μL in 1X PBS) were added to each 10 mL cell lysate. The reaction mixture was incubated at room temperature overnight with gentle

shaking.

2.5 PROTEIN MANIPULATION

2.5.1 SDS-Polyacrylamide Gel Electrophoresis of Proteins

Protein samples were separated electrophoretically on 10% (w/v) SDS-polyacrylamide gels using a Bio-Rad Mini-PROTEAN® II electrophoresis cell. Resolving gels were made by combining 1.25 mL 30% (w/v) acrylamide:0.8% (w/v) bisacrylamide (Bio-Rad); 450 μL 3 M Tris, pH 8.8; 37.5 μL 10% (w/v) SDS; 1.8 mL distilled H_2O, 1.875 μL TEMED and 187.5 μL 1.5% (w/v) APS. The gel was poured to a height that allowed space for a stacking gel of 1 cm. Immediately after pouring, isopropanol was layered on top of the acrylamide solution and the gel allowed to polymerize for 45 min to 1 h. The isopropanol was poured off, the top of the gel rinsed well with distilled H_2O and excess H_2O removed by gently blotting with Whatman 3 MM paper. A stacking gel was made by combining 0.8 mL 30% (w/v) acrylamide:0.8% (w/v) bisacrylamide; 1.25 mL 0.5 M Tris, pH 6.8; 50 μL 10% (w/v) SDS; 2.675 mL distilled H_2O, 4 μL TEMED and 250 μL 1.5% (w/v) APS. The comb was inserted, the stacking gel poured and allowed to polymerize 30 min. The wells were rinsed with 1X running buffer (25 mM Tris; 0.192 M glycine; 0.1% [w/v] SDS) prior to loading.

Samples were prepared by adding one-third volume of loading buffer (125 mM Tris, pH 6.8; 4% [w/v] SDS; 20% [v/v] glycerol; 10% [w/v] β-mercaptoethanol, 0.02% [w/v] bromphenol blue) and heating at 100°C for 5 min. For analyzing whole-cell samples, cells from 1 mL of liquid culture were harvested by microcentrifugation and resuspended in 40

µL of 100 mM DTT; 100 mM Na_2CO_3 prior to adding loading buffer. The gels were subjected to electrophoresis in 1X running buffer at room temperature at 100 V until the bromphenol blue reached the resolving gel, after which the voltage was increased to 200 V. Kaleidoscope prestained molecular weight standards (Bio-Rad), included on each gel, were treated identically to protein samples.

2.5.2 Staining of Proteins in Polyacrylamide Gels

After electrophoresis, protein gels were stained with Coomassie Blue in order to visualize polypeptides. Gels were fixed in a destaining solution of 30% (v/v) methanol, 10% (v/v) acetic acid for 15 min to 1 h to remove SDS. Gels were stained in a solution of 0.25% (w/v) Coomassie Brilliant Blue R250, 45% (v/v) methanol, 10% (v/v) acetic acid for 10 min. The staining solution was removed and the gel destained by incubation in fresh destaining solution for 2-15 h.

2.6 CYANOBACTERIAL GENE INACTIVATION

2.6.1 *In vitro* Inactivation of *crhA*

Plasmid pWM75 (Table 2.3), containing the entire 4 kb *Anabaena variabilis* UTCC 387 *Xba*I genomic fragment, was digested with *Hin*cII to yield a 2418 bp restriction fragment that contains the entire *crhA* open reading frame (ORF) plus an additional 886 bp at the 5' end and 254 bp at the 3' end. This fragment was purified from an agarose gel (Section 2.3.2), and ligated into *Hin*cII digested pBluescript KS+ (Stratagene), creating the plasmid pWM753 (Table 2.3). This plasmid contains two *Acc*I sites, located 71 bp apart, within the ORF.

pWM753 was digested with *Acc*I to release the 71 bp fragment, leaving a DNA fragment of 5308 bp (2347 bp insert plus 2961 bp vector) with cohesive ends. These were converted to blunt ends by a fill-in reaction: 0.1 mM of each of the four dNTPs and 2U Klenow fragment of DNA polymerase were added and the reaction incubated at room temperature for 30 min. The reaction was stopped by heating at 80°C for 15 min and the DNA purified on an agarose gel (Section 2.3.2). The DNA was then ligated to the spectinomycin resistance cassette, located on a 2.0 kb *Sma*I fragment isolated from pHP45 Ω (Prentki and Krisch, 1984), creating the plasmid pKSINAC (Table 2.3).

The resulting gene inactivation construct (2347 bp *Anabaena* DNA plus 2.0 kb Spectinomycin resistance cassette) was transferred to pBR322, as pBluescript KS+ lacks the *bom* site, required for plasmid transfer via conjugation. pKSINAC was digested with *Hinc*II, the 4347 bp insert isolated by gel purification (Section 2.3.2), and ligated into *Eco*RV cut pBR322, creating plasmid pBRINAC (Table 2.3). The cloning strategy of pBRINAC was confirmed by restriction analysis.

2.6.2 *In vivo* Inactivation of *crhA*

crhA was inactivated *in vivo* through homologous recombination between the wild type genomic copy and the inactivated pBRINAC copy, which was introduced into *A. variabilis* UTCC 387 by triparental mating as described by Golden and Wiest (1988). pBRINAC was transformed into *E. coli* HB101 (pRL623), creating the cargo strain HB101 (pRL623; pBRINAC). pRL623 is a helper plasmid encoding chloramphenicol resistance and M. *Ava*I, M. *Eco*47II and M. *Eco*T22I, which methylate *Ava*I, *Ava*II, and *Ava*III sites

respectively, and was obtained from J. Elhai and P. Wolk, Michigan State University. The conjugal strain used was *E. coli* JE2571 (RP4) (kindly provided by Dr. L. Frost). To set up the mating, overnight cultures of each bacterial strain were diluted 1:100 into 5 mL fresh LB broth supplemented with the appropriate antibiotics and grown to mid-log phase at 37°C on a rotary shaker. Two and one half millilitres of each culture were pelleted and resuspended in 0.5 mL LB broth. The bacteria were mixed and left at room temperature while the cyanobacteria were prepared. 100 mL of *A. variabilis* UTCC 387, grown to mid-logarithmic phase, were harvested by centrifugation at 8500 rpm in a Beckman JA-20 rotor at room temperature for 10 min and resuspended in 5 mL BG-11. Bacteria and cyanobacteria were mixed and incubated at 30°C overnight in continuous light without shaking. Aliquots of 300 μL were plated onto BG-11 plates containing spectinomycin and streptomycin at 2 μg/mL each, and incubated at 30°C in continuous light for 12-19 days. Any smeared areas of growth were re-streaked to isolate single colonies. Colonies were then innoculated onto BG-11 plates containing ampicillin (2 μg/mL) or spectinomycin and streptomycin to ensure that only ampcillin sensitive, spectinomycin/streptomycin resistant colonies were selected for further study.

CHAPTER THREE
RESULTS

3.1 ISOLATION OF AN RNA HELICASE FROM *Anabaena variabilis* UTCC 387

The isolation of RNA helicase genes in *A. variabilis* UTCC 387 was initiated during my undergraduate thesis project (Magee, 1995). In this work, PCR was used to identify segments of *A. variabilis* UTCC 387 genomic DNA that have similarity to published RNA helicase sequences. PCR reactions utilized degenerate primers based on highly conserved amino acid motifs found in these proteins. By this method, two distinct fragments, designated gwo4 and gwo29, were generated using primers corresponding to the amino acid sequences PTRELA and VLDEAD. gwo29 was unique in that it contained a FAT motif, a variation of the SAT motif; the former has not been reported in the literature. For this reason, it was hypothesized that this fragment may represent a novel RNA helicase and therefore was chosen for further study. The 780 bp nucleotide sequence and deduced amino acid sequence of gwo29 is shown in the Appendix. gwo4 was not used in this study.

In an effort to identify genomic fragments that contained the entire novel RNA helicase gene represented by gwo29, genomic DNA was isolated and digested separately with *Cla*I, *Eco*RI, *Hin*dIII and *Xba*I, electrophoresed, and transferred to a nylon membrane. The membrane was probed with the ^{32}P-labelled gwo29 fragment. The resulting hybridization pattern revealed a single hybridizing band in each lane (Magee, 1995). With the exception of the 4.4 kb *Xba*I band however, all other signals were either at or above 9.6 kb in size.

In an attempt to isolate the 4.4 kb *Xba*I fragment of *A. variabilis* UTCC 387 genomic

DNA, four separate, size-fractionated *Xba*I genomic clone banks were constructed in pBluescript KS+ (by Dr. D. Chamot). DNA from each of these clone banks (consisting of a mixture of plasmids with different insert sizes) was digested with *Xba*I and subjected to Southern hybridization analysis at high stringency, using the gwo29 partial RNA helicase gene as probe. Figure 3.1 shows a photograph of the gel before transfer (Figure 3.1A) and its resulting autoradiogram (Figure 3.1B). On the agarose gel (Figure 3.1A), the 3 kb band in each lane is the vector, pBluescript KS+. On the autoradiogram (Figure 3.1B), two distinct hybridizing bands can be seen, the sizes of which are approximately 4 kb and 4.4 kb. Clone bank 1 (lane 1) shows very faint hybridization to the 4 kb band (not visible on the photograph but apparent on overexposed autoradiograms) , clone bank 2 (lane 2) shows both bands, and clone banks 3 and 4 (lanes 3 and 4) show the 4.4 kb and 4 kb hybridizing bands, respectively. This result was unexpected as the original genomic Southern showed only a single hybridizing band. At this point it was unknown whether these two hybridizing signals represented a single gene that had been split by *Xba*I digestion, or if they were in fact two separate, but related RNA helicase genes.

Purified DNA from clone banks 3 and 4, chosen because each contained one but not both of the hybridizing fragments, was re-transformed into *E. coli* DH5α and individual colonies screened for the presence of the 4 kb and 4.4 kb inserts. Approximately 1200 colonies were screened, 600 from each clone bank, by *in situ* hybridization using the gwo29 fragment as a probe. Three separate hybridizing colonies were isolated, one from clone bank 3 and two from clone bank 4. Restriction analysis of each clone using *Xba*I indicated that the insert size in each corresponded to the sizes seen on the clone bank Southern in Figure

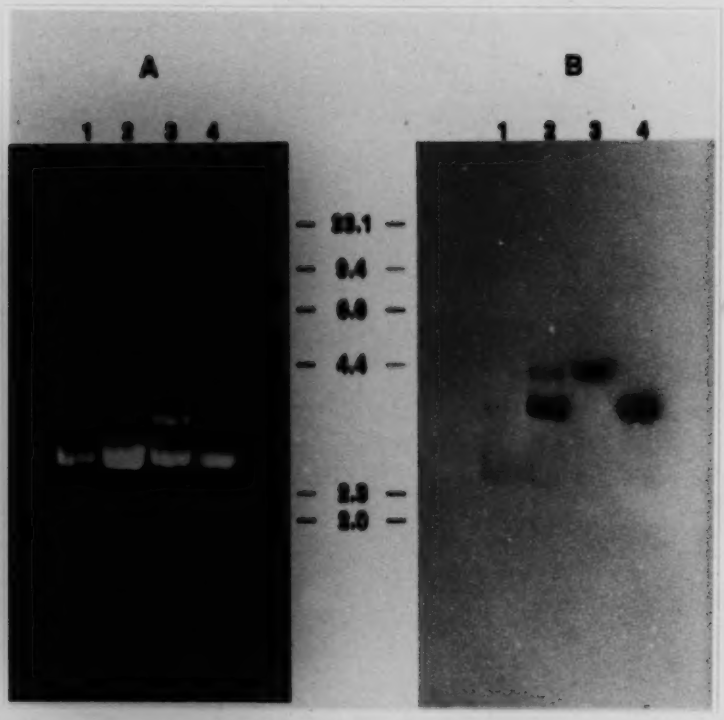


Figure 3.1. Southern analysis of *Anabaena variabilis* UTCC 387 *Xba*I genomic clone banks. A: Ethidium bromide stained agarose gel of plasmid DNA (4 μg each) from clone bank 1, clone bank 2, clone bank 3 and clone bank 4 (lanes 1-4, respectively) digested with *Xba*I. **B:** Southern analysis of the gel shown in **(A)** probed with the gwo29 fragment. Size markers (*Hin*dIII digested lambda DNA) are indicated in kb.

3.1 (data not shown).

Subsequent growth of the clone bank 3 isolate in liquid culture resulted in a loss of the insert. Attempts to re-isolate this clone by plating for single colonies and selecting for those with inserts failed, as did a repeated *in situ* hybridization experiment. As a result, further work on this isolate was abandoned and efforts focussed on the clone bank 4 isolates. Restriction analysis revealed that the two hybridizing clones isolated from clone bank 4 had identical 4 kb inserts, and it was therefore assumed that they themselves were identical. Only one, designated pWM75, was thus chosen for further analysis.

To ensure that the 4 kb insert of pWM75 originated from *A. variabilis* UTCC 387 genomic DNA and not contaminating *E. coli* DH5α genomic DNA, Southern analysis was performed on *Xba*I digests of each type of DNA, using the pWM75 insert as probe. These results are shown in Figure 3.2. Under the high stringency conditions employed, only a single hybridizing band can be seen in *A. variabilis* UTCC 387 DNA (lane 1). No signal was seen in *E. coli* DH5α DNA (lane 2), even after prolonged exposure (data not shown).

3.2 MAPPING OF pWM75

Intact pWM75 was digested individually with several restriction endonucleases. The majority of the enzymes used did not cut within the insert. These include *Bam*HI, *Bcl*I, *Bgl*II, *Bst*BI, *Bst*EII, *Cla*I, *Eco*RI, *Hin*dIII, *Nde*I, *Pst*I, *Pvu*II, *Sac*I, *Sac*II, *Sal*I and *Xho*I. Enzymes that were able to cleave the insert were *Acc*I, *Eco*RV and *Hin*cII. Based on single and double digests of pWM75 by these three enzymes, a restriction map, shown in Figure 3.3A, was constructed. Only the insert DNA is shown, flanked by *Xba*I ends. It is important

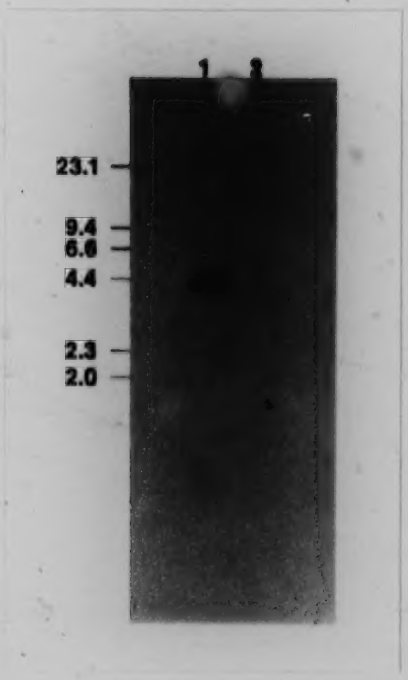
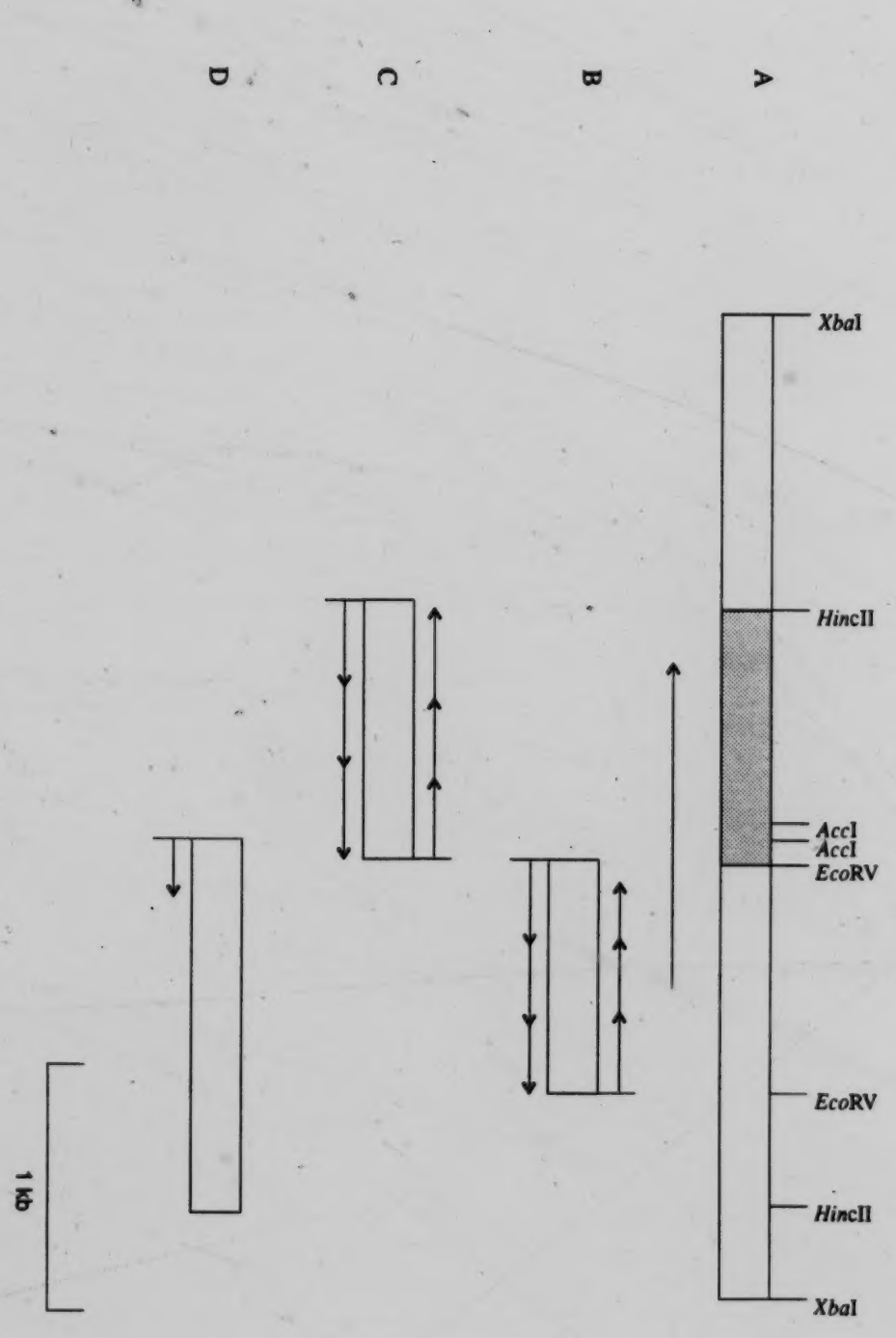


Figure 3.2. Southern hybridization of *A. variabilis* UTCC 387 and *E. coli* DH5α genomic DNAs with pWM75. *A. variabilis* UTCC 387 (lane 1) and *E. coli* DH5α (lane 2) genomic DNAs (1.5 μg each) were digested with *Xba*I, separated by electrophoresis and subjected to Southern transfer. The membrane was probed with the 4 kb *Xba*I insert of pWM75. Size markers (*Hin*dIII digested lambda DNA) are indicated in kb.

Figure 3.3. Restriction map and sequencing strategy of pWM75. A: Restriction map of the pWM75 insert. Restriction endonucleases which cleave within the insert include *Hinc*II, *Acc*I and *Eco*RV. The smallest fragment to which gwo29 hybridized is shaded. The arrow underneath the map indicates the open reading frame and is to scale. **B-D:** Arrows indicate the direction of sequencing on various subclones. Subclones include: *Eco*RV/*Eco*RV of pWM752 **(B)**, *Hinc*II/*Eco*RV of pWM754 **(C)**, and *Acc*I/*Hinc*II of pWM755 **(D)**. The scale is indicated for 1 kb.


A
B
C
D
XbaI
HincII
AccI
AccI
EcoRV
EcoRV
HincII
XbaI
1kb

to note that although two *Acc*I sites appear on the map, only one was determined by restriction mapping; the other was found during the course of sequencing (Section 3.3).

Single and double digests of pWM75 DNA by *Acc*I, *Eco*RV, and *Hin*cII were separated by electrophoresis (Figure 3.4A), subjected to a Southern transfer and probed with the gwo29 fragment. Autoradiography (Figure 3.4B) revealed a number of hybridizing fragments. The smallest single region which hybridized to the probe was an approximately 1 kb *Hin*cII/*Eco*RV fragment internal to the pWM75 insert (lane 8), shown as the shaded area in Figure 3.3A.

3.3 SEQUENCING OF THE pWM75 INSERT AND SEQUENCE ANALYSIS

From the hybridization results in Figure 3.4, it appeared that the putative RNA helicase gene was contained entirely within the 4 kb *Xba*I genomic DNA fragment, as the smallest hybridizing region was an approximately 1 kb area near the centre of the pWM75 insert. This hypothesis was confirmed by partial sequencing of the ends of the pWM75 insert. Subjecting the sequenced regions to a homology search (using the BLAST program of the Genetics Computer Group (GCG) Sequence Analysis Software package version 8.1) revealed no significant homology to any gene, and more importantly to no known RNA helicase (data not shown). If the clone bank Southern (Figure 3.1B) was showing a single gene split by *Xba*I digestion, then one of the ends of the 4 kb fragment would include part of the open reading frame of the RNA helicase.

In order to facilitate sequencing of the hybridizing region, subclones of various restriction fragments were constructed in pBluescript KS+. Shown in Figure 3.3 parts B-D,

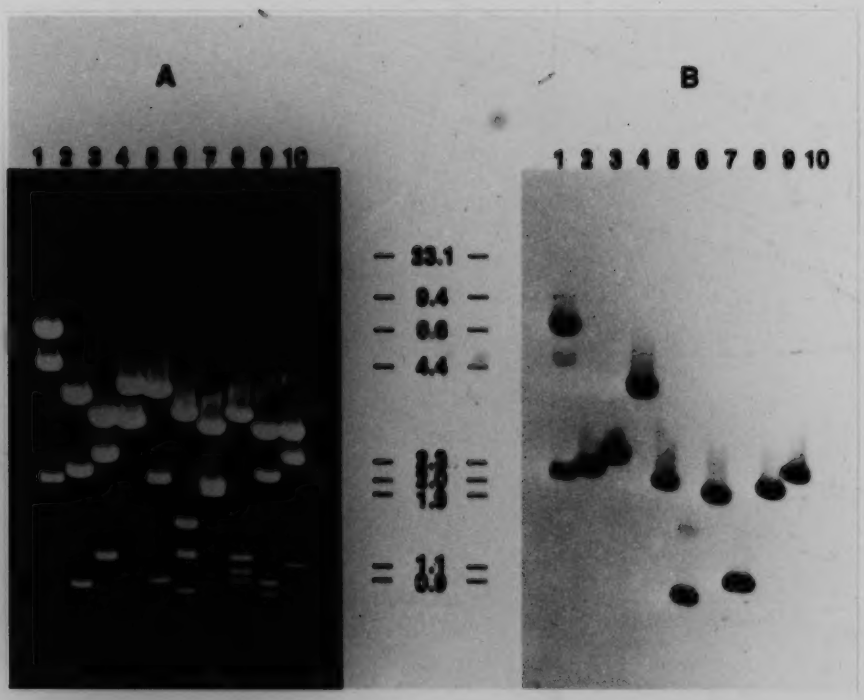


Figure 3.4. Southern analysis of pWM75. A: Ethidium bromide stained gel showing restriction endonuclease digests of pWM75 after electrophoresis. Digests include: *Acc*I (lane 1), *Eco*RV (lane 2), *Hin*cII (lane 3), *Xba*I (lane 4), *Acc*I/*Eco*RV (lane 5), *Acc*I/*Hin*cII (lane 6), *Acc*I/*Xba*I (lane 7), *Eco*RV/*Hin*cII (lane 8), *Eco*RV/*Xba*I (lane 9), and *Hin*cII/*Xba*I (lane 10). **B:** The gel in **(A)** was subjected to Southern transfer and probed with the gwo29 fragment. Size markers (*Hin*dIII digested lambda DNA and *Bst*NI digested pBR322 DNA) are indicated in kb.

these subclones were: pWM752, a 936 bp *Eco*RV/*Eco*RV fragment (B); pWM754, a 1028 bp *Hinc*II/*Eco*RV fragment (C); and pWM755, a 1487 bp *Acc*I/*Hinc*II fragment (D). The identity of each subclone was confirmed by determining the size of each insert by restriction analysis. Both strands of pWM752 and pWM754 were sequenced, as shown by arrows in Figure 3.3B and 3.3C. Oligonucleotide primers were designed and used in order to extend the sequence in the direction of the arrows. Each arrow in Figure 3.3 indicates the sequence generated by a specific primer. pWM755 (Figure 3.3D) was used to confirm the sequence of the adjoining ends of pWM752 and pWM754, as these latter two plasmids do not overlap.

The nucleotide and deduced amino acid sequences of the entire 1964 bp *Hinc*II/*Eco*RV region of pWM75 is shown in Figure 3.5. This region encodes an open reading frame of 1278 bp capable of coding for 425 amino acids. The predicted molecular weight of this protein was determined to be 47 kDa by the GCG PEPTIDESORT program. The direction of transcription is indicated by the arrow below the pWM75 restriction map (Figure 3.3A).

Analysis of the deduced amino acid sequence revealed that a DEAD motif is present, suggesting that this protein is a member of the DEAD box family of RNA helicases. Indeed, all nine of the highly conserved motifs found in this family are present and are indicated in bold in Figure 3.5. Interestingly, the SAT motif has been replaced by FAT, suggesting that a serine to phenylalanine change at the first position has occurred at some point in this organism. This alteration is also seen in the PCR generated fragment, gwo29 (see Appendix).

The gene encoding this putative RNA helicase was named *crhA*, for *c*yanobacterial

Figure 3.5. Nucleotide sequence of the *Hin*cII/*Eco*RV region of the pWM75 insert. The sequence extends from the 5' *Eco*RV site to the 3' *Hin*cII site shown in Figure 3.3A. Numbering is indicated relative to the adenine in the start codon of the open reading frame (+1). Putative -10 promoter regions are underlined and a possible Shine-Dalgarno sequence is double underlined. A putative rho-dependent transcriptional stop stem-loop structure is indicated by inverted arrows. The deduced amino acid sequence of the open reading frame is shown below the DNA sequence. Conserved DEAD box RNA helicase amino acid sequence motifs are shown in bold.

```
-429 GATATCATCTCCCTGCTTACCTATACTTTAGGCTTCGGTTTAACTCAGTATGCAATCGCC -370
-369 ACTAGCACCAGACGACTAGTTAGCGATAGTCTATCCACCATTGTTCGTTTTGTAGGTTTT -310
-311 GCTTTTATAGCGATCGGTTTTGTATTTTGCGGTAACTTCATCAATTTTTTAGGGGCTGGT -250
-249 AATTTTTAACATATCTCACGGGGTGCAATCTTCGCGCCCCTACTAGTCCATCGAATCGTC -190
-189 ATTTCCAACTATTAATATTAAAGTTTAGAGAAATTGGATTATATGTAACCTGTACTCTGT -130
-129 TAAGATTCACCATTGGGGTATTCGCTATCAGTCTTGGCGCTACTGCCCATCCCGCCCCTC -70
 -69 AAACCTTTGTCCGTCCGCCTAAGACTGATACCGCTACTGGTGACAGGCCGATGTTATATC -10

  -9 TGGAGTTCTATGTCTTTTTCTCATCTCGGCTTGTCCAATGAAATTATCAATGCTGTTACT 51
   1        M  S  F  S  H  L  G  L  S  N  E  I  I  N  A  V  T  17

  52 GAGTTGGGGTACACCAAACCCACACCCATCCAGATGCAGTCTATTCCTGCTGTCTTATCA 111
  18 E  L  G  Y  T  K  P  T  P  I  Q  M  Q  S  I  P  A  V  L  S  37

 112 GGACGAGATTTGCTACGTGGCGCTCAAACTGGAACTGGGAAAACTGCCAGCTTTACTCTC 171
  38 G  R  D  L  L  R  G  A  Q  T  G  T  G  K  T  A  S  F  T  L  57

 172 CCCCTCCTACATTATTATCCCAAGATAGTCTTAAAGGCCCTCCAATGCTTCCTCACCATT 231
  58 P  L  L  H  Y  Y  P  K  I  V  L  K  A  L  Q  C  F  L  T  I  77

 232 CGGGCGCTAATTCTTACCCCGACTCGTGAACTCGCCGCACAGGTGGAGTCAAGCGTGCGT 291
  78 R  A  L  I  L  T  P  T  R  E  L  A  A  Q  V  E  S  S  V  R  97

 292 GACTACGGCAAGTACTTGAAGCTGAACTCAATGGTGATGTTTGGTGGAGTCAGCATTAAT 351
  98 D  Y  G  K  Y  L  K  L  N  S  M  V  M  F  G  G  V  S  I  N  117

 352 CCCCAAAAACAGCGTTTAAAGGGTCGTGTAGATATTCTGGTTGCTACCCCAGGGCGACTG 411
 118 P  Q  K  Q  R  L  K  G  R  V  D  I  L  V  A  T  P  G  R  L  137

 412 CTAGACCATGTGCAGCAGGGAACGGTAAACCTTTCACAGATTGAGATTCTGGTGCTGGAT 471
 138 L  D  H  V  Q  Q  G  T  V  N  L  S  Q  I  E  I  L  V  L  D  157

 472 GAAGCCGACCGGATGTTGGATATGGGCTTTATTCGTGATATCCGTCGTATCCTCTCCCTG 531
 158 E  A  D  R  M  L  D  M  G  F  I  R  D  I  R  R  I  L  S  L  177

 532 TTGCCCAAACAGCGACAAAACTTGCTATTCTTCGCTACATTCTCAGACAAAATCAAGGAA 591
 178 L  P  K  Q  R  Q  N  L  L  F  F  A  T  F  S  D  K  I  K  E  197

 592 CTCGCCGCCGGTCTACTGAATCGCCCGCAGATGATCGAAGTAGCACGCCGTAACGTTACA 651
 198 L  A  A  G  L  L  N  R  P  Q  M  I  E  V  A  R  R  N  V  T  217

 652 GCCGATACTGTGACACAAAAAGTCTACAAAATAGAGCGCGATAGAAAACGAGATTTACTG 711
 218 A  D  T  V  T  Q  K  V  Y  K  I  E  R  D  R  K  R  D  L  L  237

 712 GCTCACCTGATTCGGAAAGATAATTGGTATCAAGTACTAGTATTTACTCGTACCAAGTAT 771
 238 A  H  L  I  R  K  D  N  W  Y  Q  V  L  V  F  T  R  T  K  Y  257

 772 GGTGCTGACCGTCTAGTTAAACAATTAGGCCATGAGCGCATTCAAGCTCTAGCTATCCAC 831
 258 G  A  D  R  L  V  K  Q  L  G  H  E  R  I  Q  A  L  A  I  H  277

 832 GGTAATAAGAGCCAGTCGGCGCGTACCCACGCTCTGGCAAAGTTCAAAAATGGTAGTTTA 891
 278 G  N  K  S  Q  S  A  R  T  H  A  L  A  K  F  K  N  G  S  L  297

 892 CAAGTTTTAGTAGCAACTGACATTGCTGCACGAGGACTGGACATCAGCGAATTGCCTTAT 951
 298 Q  V  L  V  A  T  D  I  A  A  R  G  L  D  I  S  E  L  P  Y  317

 952 GTGGTCAATTTCGATTTGCCCTATGTACCAGAAGATTATGTTCATCGTATTGGTCGCACT 1011
 318 V  V  N  F  D  L  P  Y  V  P  E  D  Y  V  H  R  I  G  R  T  337

1012 GGTCGCGCTGGTGCATCAGGTGAGGCTGTATCGCTGGTGAGCGCCGATGAATATCATTTG 1071
 338 G  R  A  G  A  S  G  E  A  V  S  L  V  S  A  D  E  Y  H  L  357

1072 TTGGCAGATATTGAAAAACTGATTGAAAAGCGATTGCCTTTTGAATTGGTAGCGGGTATT 1131
 358 L  A  D  I  E  K  L  I  E  K  R  L  P  F  E  L  V  A  G  I  377
```

```
1132 GGAGCTAATTCCCAAGCTAAACCCGAACCAACTCAGGATGAACGCAAGCAAAAACCCAAA 1191
 378 G  A  N  S  Q  A  K  P  E  P  T  Q  D  E  R  K  Q  K  P  K   397

1192 GATAGTCAGCATCAGCCTCGCTCTGCTGCTCAAGGTGTGCCAAAGAAATCAGGGAAAAAA 1251
 398 D  S  Q  H  Q  P  R  S  A  A  Q  G  V  P  K  K  S  G  K  K   417

1252 CGTTTAACTAATTCTGGAAAAAGGTAATACGTACAAAAGAAAGTATGAAAAACTCTACCC 1311
 418 R  L  T  N  S  G  K  R  *                                     425

1312 ACAAGGGGGTAGAGTTTTCGGTAATGGGTAATGGGTAATTTATGAGAGCCAAGATTACTC 1371
1372 AATCGTTAGTATGAGCGATCGCTTTAGAAGGCTCCGACTCTAGCGACGATAACGTAAATT 1431
1432 GTTTTACCAATCAGTAGGAGGTCGTATAGAGGATGCCAATTATTTTGATATTGCAAATCT 1491
1492 AAATCTACTATCTGTTCAAAGTCTTTCACATGAGAGCGTCCGTTATC               1538
```

*R*NA *h*elicase *A*. This name is provisional at this point, pending the elucidation of the function of the protein encoded by this gene.

The DNA sequences surrounding the open reading frame were also analyzed. A number of putative *E. coli* -10 regions, shown underlined in Figure 3.5, were found in the DNA sequence 5' to the open reading frame start codon. No consensus *E. coli* -35 region could be identified. A putative Shine-Dalgarno sequence, shown double underlined, was also associated with the presumed initiator ATG. This ATG was chosen as the start codon based on its position relative to the conserved amino acid motifs and because it was the first methionine residue in the open reading frame. The 3' untranslated region contained a potential rho-dependent terminator 21 bp downstream of the stop codon. The terminator consists of a short stem-loop structure composed of a 12 bp stem and a 6 bp loop. The inverse repeats that form the stem-loop are designated by arrows in Figure 3.5.

In order to gain insight into the possible function of *crhA*, the deduced amino acid sequence of its open reading frame was subjected to a BLAST homology search. A number of established and putative RNA helicases were identified from GenBank that have significant identity with the CrhA deduced amino acid sequence. Figure 3.6 shows an amino acid sequence alignment of CrhA with various DEAD box RNA helicases from two other Gram negative species (*E. coli* and *Synechocystis* sp. strain PCC 6803) and a yeast and a mouse gene for comparison. The nine DEAD box family conserved amino acid motifs are shown in bold. These proteins were chosen for the following reasons: the *E. coli* proteins because they are the most extensively studied prokaryotic RNA helicases, the *Synechocystis* sp. strain PCC 6803 RNA helicase because it is the only other described cyanobacterial RNA

Figure 3.6. Alignment of deduced RNA helicase amino acid sequences. Sequences were obtained from GenBank and were aligned using the PILEUP and PRETTY programs of the University of Wisconsin Genetics Computer Group Sequence Analysis Software program version 8.1. Sequences include: ChrA from *A. variabilis* UTCC 387 (this work), RhlE from *E. coli*, a putative RNA helicase from *Synechocystis* sp. strain PCC 6803 (designated 6803), CsdA, RhlB, SrmB and DbpA, all from *E. coli*, PL10 from *Mus musculus* and YHR065C from *Sacchromyces cerevisiae*. Accession numbers can be obtained from Table 3.1. Highly conserved DEAD box family amino acid sequence motifs are shown in bold.

```
   ChrA  .......... .......... .......... .......... ..........
   RhlE  .......... .......... .......... .......... ..........
   6803  .......... .......... .......... .......... ..........
   CsdA  .......... .......... .......... .......... ..........
   RhlB  .......... .......... .......... .......... ..........
   SrmB  .......... .......... .......... .......... ..........
   DbpA  .......... .......... .......... .......... ..........
   PL10  MSHVAEEDEL GLDQQLAGLD LTSRDSQSGG STASKGRYIP PHLRNREAAK
YHR065C  .......... .......... .......... .......... ..........

   ChrA  .......... .......... .......... .......... ..........
   RhlE  .......... .......... .......... .......... ..........
   6803  .......... .......... .......... .......... ..........
   CsdA  .......... .......... .......... .......... ..........
   RhlB  .......... .......... .......... .......... ..........
   SrmB  .......... .......... .......... .......... ..........
   DbpA  .......... .......... .......... .......... ..........
   PL10  AFYDKDGSRW SKDKDAYSSF GSRSDTRAKS SFFSDRGGSG SRGRFDERGR
YHR065C  ......MKNL KLINEHLAIV HGIHRIVTYL RSNIYQDSVK YFTESAIKMS

   ChrA  .......... .......... .......... .......... ..........
   RhlE  .......... .......... .......... .......... ..........
   6803  .......... .......... .......... .......... ..........
   CsdA  .......... .......... .......... .......... ..........
   RhlB  .......... .......... .......... .......... ..........
   SrmB  .......... .......... .......... .......... ..........
   DbpA  .......... .......... .......... .......... ..........
   PL10  SDYESVGSRG GRSGFGKFER GGNSRWCDKA DEDDWSKPLP PSERLEQELF
YHR065C  KIVKRKEKKA NDELTSLAEK IRAKALENQK KLIEAEKEGG SESDSEEDAT

   ChrA  .......... .......... ........MS FSHLG.LSNE IINAVTELGY
   RhlE  .......... .......... ........MS FDSLG.LSPD ILRAVAEQGY
   6803  .......... .......... ..MTNTLTST FADLG.LSEK RCQLLADIGF
   CsdA  ......MMSY VDWPPLILRH TYYMAEFETT FADLG.LKAP ILEALNDLGY
   RhlB  .......... .......... MSKTHLTEQK FSDFA.LHPK VVEALEKKGF
   SrmB  .......... .......... ....MTVTT FSELE.LDES LLEALQDKGF
   DbpA  .......... .......... .......MTA FSTLNVLPPA QLTNLNELGY
   PL10  SGGNTGINFE KYDDIPVEAT GNNCPPIES FSDVE.MGEI IMGNIELTRY
YHR065C  AEKKKVLKSK SKSTVSTQNE NTNEDESFLS FSELN.LVPE LIQACKNLNY

   ChrA  TKPTPIQMQS IPAVLSGRDL LRGAQTGTGK TASFTLPLL. HYYPKIVLKA
   RhlE  REPTPIQQQA IPAVLEGRDL MASAQTGTGK TAGFTLPLL. QHLITRQPHA
   6803  EAPTQIQTEA IPLLLSGRDM LAQSQTGTGK TAAFALPLM. DRID......
   CsdA  EKPSPIQAEC IPHLLNGRDV LGMAQTGSGK TAAFSLPLL. QNLD......
   RhlB  HNCTPIQALA LPLTLAGRDV AGQAQTGTGK TMAFLTSTF. HYLLSHPAIA
   SrmB  TRPTAIQAAA IPPALDGRDV LGSAPTGTGK TAAYLLPAL. QHLLDFP...
   DbpA  LTMTPVQAAA LPAILAGKDV RVQAKTGSGK TAAFGLGLL. QQIDA.....
   PL10  TRPTPVQKHA IPIIKEKRDL MACAQTGSGK TAAFLLPILS QIYTDGPGEA
YHR065C  SKPTPIQSKA IPPALEGHDI IGLAQTGSGK TAAFAIPILN RLWHDQEPYY
```

```
ChrA    LQCFLT.... ......IRAL ILTPTRELAA QVESSVRDYG KYLK.LNSMV
RhlE    .KGRRP.... ......VRAL ILTPTRELAA QIGENVRDYS KYLN.IRSLV
6803    .PE.GD.... ......LQAL ILTPTRELAQ QVAEAMKDFS .HERRLFILN
CsdA    .PELKA.... ......PQIL VLAPTRELAV QVAEAMTDFS KHMRGVNVVA
RhlB    DRKVNQ.... ......PRAL IMAPTRELAV QIHADAEPLA E.ATGLKLGL
SrmB    RKKSGP.... ......PRIL ILTPTRELAM QVSDHARELA KHTH.LDIAT
DbpA    ..SLFQ.... ......TQAL VLCPTRELAD QVAGELRRLA RFLPNTKILT
PL10    LRAMKENGKY GRRKQYPISL VLAPTRELAV QIYEEARKFS .YRSRVRPCV
YHR065C .......... ........AC ILAPTRELAQ QIKETFDSLG SLM.GVRSTC

ChrA    MFGGVSINPQ KQRLKGRVDI LVATPGRLLD HVQQGT.VNL SQIEILVLDE
RhlE    VFGGVSINPQ MMKLRGGVDV LVATPGRLLD LEHQNA.VKL DQVEILVLDE
6803    VYGGQSIERQ IRSLERGVQI VVGTPGRVID LIDRKK.LKL ETIQWVVLDE
CsdA    LYGGQRYDVQ LRALRQGPQI VVGTPGRLLD HLKRGT.LDL SKLSGLVLDE
RhlB    AYGGDGYDKQ LKVLESGVDI LIGTTGRLID YAKQNH.INL GAIQVVVLDE
SrmB    ITGGVAYMNH AEVFSENQDI VVATTGRLLQ YIKEEN.FDC RAVETLILDE
DbpA    LCGGQPFGMQ RDSLQHAPHI IVATPGRLLD HLQKGT.VSL DALNTLVMDE
PL10    VYGGADIGQQ IRDLERGCHL LVATPGRLVD MMERGK.IGL DFCKYLVLDE
YHR065C IVGGMNMMDQ ARDLMRKPHI IIATPGRLMD HLENTKGFSL RKLKFLVMDE

ChrA    ADRMLDMGFI RDIRRILSL. .LPKQRQ..N LLFFATF.SD KIKELAAGLL
RhlE    ADRMLDMGFI HDIRRVLTK. .LPAKRQ..N LLFSATF.SD DIKALAEKLL
6803    ADEMLSMGFI DDVKTILRK. .TPPTRQ..T ACFSATM.PR EIKELVNQFL
CsdA    ADEMLRMGFI EDVETIMAQ. .IPEGHQ..T ALFSATM.PE AIRRITRRFM
RhlB    ADRMYDLGFI KDIRWLFRR. .MPPANQRLN MLFSATL.SY RVRELAFEQM
SrmB    ADRMLDMGFA QDIEHIAGE. .TRWRKQ..T LLFSATLEGD AIQDFAERLL
DbpA    ADRMLDMGFS DAIDDVIRF. .APASRQ..T LLFSATW.PE AIAAISGRVQ
PL10    ADRMLDMGFE PQIRRIVEQD TMPPKGVRHT MMFSATF.PK EIQMLARDFL
YHR065C ADRLLDMEFG PVLDRILK.. .IIPTQERTT YLFSATMTS. KIDKLQRASL

ChrA    NRPQMIEVAR RNVTADTVTQ KVYKIERD.R KRDLLAHLIR KDNWYQVLVF
RhlE    HNPLEIEVAR RNTASDQVTQ HVHFVDKK.R KRELLSHMIG KGNWQQVLVF
6803    NDPALVTVKQ TQSTPTRIEQ QLYHVPRGWS KAKALQPILE MEDPESAIIF
CsdA    KEPQEVRIQS SVTTRPDISQ SYWTV.WGMR KNEALVRFLE AEDFDAAIIF
RhlB    NNAEYIEVEP EQKTGHRIKE ELFYPSNE.E KMRLLQTLIE EEWPDRAIIF
SrmB    EDPVEVSANP STRERKKIHQ WYYRADDLEH KTALLVHLLK QPEATRSIVF
DbpA    RDPLAIEIDS TDALPP.IEQ QFYETSSK.G KIPLLQRLLS LHQPSSCVVF
PL10    DEYIFLAVGR VGSTSENITQ KVVWVEEADK RSFLLDLLNA TGKDSLILVF
YHR065C TNPVKCAVSN KYQTVDTLVQ TLMVVPGGL. KNTYLIYLLN EFIGKTMIIF

ChrA    TRTKYGADRL VKQLGHERIQ ALAIHGNKSQ SARTHALAKF KNGSLQVLVA
RhlE    TRTKHGANHL AEQLNKDGIR SAAIHGNKSQ GARTRALADF KSGDIRVLVA
6803    VRTKQTAADL TSRLQEAGHS VDEYHGNLSQ SQRERLVHRF RDGKIKLVVA
CsdA    VRTKNATLEV AEALERNGYN SAALNGDMNQ ALREQTLERL KDGRLDILIA
RhlB    ANTKHRCEEI WGHLAADGHR VGLLTGDVAQ KKRLRILDEF TRGDLDILVA
SrmB    VRKRERVHEL ANWLREAGIN NCYLEGEMVQ GKRNEAIKRL TEGRVNVLVA
DbpA    CNTKKDCQAV CDALNEVGQS ALSLHGDLEQ RDRDQTLVRF ANGSARVLVA
PL10    VETKKGADSL EDFLYHEGYA CTSIHGDRSQ RDREEALHQF RSGKSPILVA
YHR065C TRTKANAERL SGLCNLLEPS ATALHGDLNQ NQRMGSLDLF KAGKRSILVA
```

```
   ChrA  TDIAARGLDI SELPYVVNFD LPYVPEDYVH RIGRTGRAGA SGEAVSLVSA
   RhlE  TDIAARGLDI EELPHVVNYE LPNVPEDYVH RIGRTGRAAA TGEALSLVCV
   6803  TDIAARGLDV NNLSHVVNFD LPDNAETYIH RIGRTGRAGK TGKAIALVEP
   CsdA  TDVAARGLDV ERISLVVNYD IPMDSESYVH RIGRTGRAGR AGRALLFVEN
   RhlB  TDVAARGLHI PAVTHVFNYD LPDDCEDYVH RIGRTGRAGA SGHSISLACE
   SrmB  TDVAARGIDI PDVSHVFNFD MPRSGDTYLH RIGRTARAGR KGTAISLVEA
   DbpA  TDVAARGLDI KSLELVVNFE LAWDPEVHVH RIGRTARAGN SGLAISFCAP
   PL10  TAVAARGLDI SNVKHVINFD LPSDIEEYVH RIGRTGRVGN LGLATSFFNE
YHR065C  TDVAARGLDI PSVDIVVNYD IPVDSKSYIH RVGRTARAGR SGKSISLVSQ

   ChrA  DEYHLLADIE KLI.......  .....EKRLP FELVAGIGAN SQAKPEP...
   RhlE  DEHKLLRDIE KLL.......  .....KKEIP RIAIPGYEPD PSIKAEP...
   6803  IDRRLLRSIE NRL.......  .....KQQIE VCTIPNRSQV EAKRIEK...
   CsdA  RERRLLRNIE RTM.......  .....KLTIP EVELPNAELL GKRRLEK...
   RhlB  EYALNLPAIE TYI.......  .....GHSIP VSKY.....N PDALMTD...
   SrmB  HDHLLLGKVG RYI.......  .....EEPIK ARVIDELRPK TRAPSEK...
   DbpA  EEAQRANIIS DML.......  .....QIKLN WQTPPANSSI ATLEAEMATL
   PL10  RNINITKDLL DLLVEAKQEV PSWLENMAFE HHYKGGSRGR SKSRFSGGFG
YHR065C  YDLELILRIE EVL.......  GKKLPKESVD KNIILTLRDS VDKANGEVVM

   ChrA  .TQDERKQKP KDSQHQP.RS AAQGVPKKSG KKRLTNSGKR ..........
   RhlE  .IQNGRQQRG GGGRGQG.GG RGQQQPRRGE GGAKSASAKP AEKPSRRLGD
   6803  .LQEQLKEAL TGERMASFLP LVREL.....  SDEYDAQAIA AAALQM....
   CsdA  .FAAKVQQQL ESSDLDQYRA LLSKIQPTAE GEELDLETLA AALLKMAQGE
   RhlB  .LPKPLRLTR PRTGNGPRRT GAPRNRRRSG ..........  ..........
   SrmB  .QTGKPSKKV LAKRAEKKKA KEKEKPRVKK RHRDTKNIGK RRKPSGTGVP
   DbpA  CIDGGKKAKM RPGDVLGALT GDIGLDGADI GKIAVHPAHV YVAVRQAVAH
   PL10  ARDYRQSSGA SSSSFSSGRA SNSRSGGGSH GSSRGFGGGS YGGFYNSDGY
YHR065C  EMNRRNKEKI ARGKGRRGRM MTRENMDMGE R.........  ..........

   ChrA  ..........  ..........  ..........  ..........  ..........
   RhlE  AKPAGEQQRR RRPRKPAAAQ  ..........  ..........  ..........
   6803  ...IYDQSCP HWMKSD....  ..........  ..........  .....WEVPE
   CsdA  RTLIVPPDGP MRPKREFRDR DDRGPRDRND RGPRGDREDR PRRERRDVGD
   RhlB  ..........  ..........  ..........  ..........  ..........
   SrmB  PQTTEE....  ..........  ..........  ..........  ..........
   DbpA  KAWKQLQGGK IKGKTCRVRL LK........  ..........  ..........
   PL10  GGNYSSQGVD WWGN......  ..........  ..........  ..........
YHR065C  ..........  ..........  ..........  ..........  ..........

   ChrA  ..........  ..........  ..........  ..........  ..........
   RhlE  ..........  ..........  ..........  ..........  ..........
   6803  VD........  ..........  ..........  ..........  ..........
   CsdA  MQLYRIEVGR DDGVEVRHIV GAIANEGDIS SRYIGNIKLF ASHSTIELPK
   RhlB  ..........  ..........  ..........  ..........  ..........
   SrmB  ..........  ..........  ..........  ..........  ..........
   DbpA  ..........  ..........  ..........  ..........  ..........
   PL10  ..........  ..........  ..........  ..........  ..........
YHR065C  ..........  ..........  ..........  ..........  ..........
```

```
   ChrA  .......... .......... .......... .......... ..........
   RhlE  .......... .......... .......... .......... ..........
   6803  .......... .....FNKPV .......... ....LRRGRN AG.......G
   CsdA  GMPGEVLQHF TRTRILNKPM NMQLLGDAQP HTGGERRGGG RGFGGERREG
   RhlB  .......... .......... .......... .......... ..........
   SrmB  .......... .......... .......... .......... ..........
   DbpA  .......... .......... .......... .......... ..........
   PL10  .......... .......... .......... .......... ..........
YHR065C  .......... .......... .......... .......... ..........

   ChrA  .......... .......... .......... .......... ....
   RhlE  .......... .......... .......... .......... ....
   6803  GQNKSGGGYQ GKPGKPRRSS GGRRPAYSDR QQ........ ....
   CsdA  GRNFSGERRE GGRGDGRRFS GERREGRAPR RDDSTGRRRF GGDA
   RhlB  .......... .......... .......... .......... ....
   SrmB  .......... .......... .......... .......... ....
   DbpA  .......... .......... .......... .......... ....
   PL10  .......... .......... .......... .......... ....
YHR065C  .......... .......... .......... .......... ....
```

helicase, and YHR065C (yeast) and PL10 (mouse) to represent lower and higher eukaryotes, respectively. The alignment shows that outside of the conserved motifs, there is no significant homology between these proteins.

Table 3.1 shows a matrix of percent identity and percent similarity among the nine proteins in Figure 3.6. Values were obtained for pairwise combinations using the GCG GAP program. Percent identity ranges from 60.6% (CrhA/RhlE) to 28.6% (SrmB/PL10) and percent similarity ranges from 74.3% (CrhA/RhlE) to 52.0% (RhlB/PL10). Overall, CrhA appears to be the most similar to the *E. coli* RNA helicase RhlE. It is important to note that all other pairwise combinations revealed percent identities of approximately 37%±3 and percent similarities of approximately 58%±5. Therefore, the results for CrhA/RhlE, whose values are much higher, are significant.

To determine the context in which *crhA* resides on the *A. variabilis* UTCC 387 genome, homology searches were performed on the sequenced regions outside of the *crhA* open reading frame. A BLAST search of the 429 bp 5' to the open reading frame revealed no significant homology to any protein in the GenBank database. In contrast, a search of the 260 bp region 3' to the open reading frame revealed a number of matches to proteins involved in carbohydrate metabolism. These include the *Streptococcus salivarius* undecaprenyl-phosphate glycosyl-l-phosphate transferase, the *S. agalactiae* galactosyl transferase, and the *S. pneumoniae* UDP-glycosyl transferase. All of these enzymes have a role in the transfer of sugar moieties. The gene that is encoded in the 260 bp region whose open reading frame is related to these transferase enzymes is transcribed in a direction opposite to the *crhA* open reading frame, suggesting the *chrA* gene is not part of an operon

Table 3.1 Comparison of Percent Identity and Percent Similarity Among Selected RNA Helicase Amino Acid Sequences[a]

	CrhA	RhlE	6803	CsdA	DbpA	PL10	SrmB	RhlB	YHR065C
CrhA		74.3	57.8	60.0	55.9	56.7	56.8	55.0	56.7
RhlE[b]	60.6		62.4	63.8	54.9	58.9	58.4	59.3	55.8
6803[c]	38.6	39.2		63.3	57.1	56.4	54.3	57.7	53.5
CsdA[d]	38.5	40.2	44.6		58.6	58.3	59.6	56.1	56.2
DbpA[e]	38.0	37.4	34.4	38.2		54.7	53.8	55.0	55.3
PL10[f]	39.0	39.1	36.4	37.1	33.6		52.6	52.0	53.0
SrmB[g]	36.8	37.6	34.7	37.4	33.7	28.6		58.2	56.0
RhlB[h]	36.3	37.7	37.7	38.6	34.9	33.5	37.9		52.3
YHR065C[i]	37.4	36.3	34.7	37.2	34.7	33.3	34.2	34.6	

[a]Percent identities and percent similarities are listed in the lower and upper parts of the table, respectively. Sequences were obtained from GenBank.

[b]From *E. coli* (Accession #P25888)

[c]From *Synechocystis* sp. strain PCC 6803 (Accession #D64004); no designation given

[d]From *E. coli* (Accession #P23304); also known as DeaD

[e]From *E. coli* (Accession #P21693)

[f]From *Mus musculus* (Accession #P16381)

[g]From *E. coli* (Accession #P21507)

[h]From *E. coli* (Accession #P24229)

[i]From *Saccharomyces cerevisiae* (Accession #P38712)

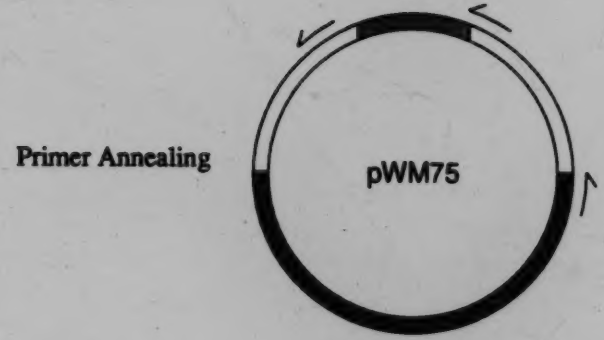
in this direction.

3.4 OVEREXPRESSION OF CrhA

crhA was identified as a putative RNA helicase on the basis of amino acid sequence similarity to known members of this class of enzyme. To fully establish this gene as a true RNA helicase, however, assays for RNA helicase activity (RNA-dependent ATPase and ATP-dependent RNA unwinding) must be performed. In an effort to start this process, attempts were made to purify the CrhA protein. The pGEX system (Pharmacia) was employed for this process, in which CrhA can be overexpressed in *E. coli* as a fusion protein from an expression vector and purified by affinity chromatography. The protein of interest can then be isolated by protease cleavage of the fusion protein and removal of the vector-encoded peptide.

In order to create an in-frame translational fusion, the DNA encoding the *crhA* open reading frame had to be isolated. As no appropriate restriction sites were present, site-directed mutagenesis was used to create a *Bam*HI site immediately preceding the initiator codon and an *Eco*RI site 100 bp downstream of the stop codon of the *crhA* open reading frame. A flowchart of both the mutagenesis procedure and the isolation of the DNA encoding the *crhA* open reading frame is shown in Figure 3.7. After its purification, the 1.4 kb *Bam*HI/*Eco*RI fragment that contained this coding region was cloned into pGEX-2T, creating plasmid pGX29. pGX29 was confirmed by restriction analysis and partial DNA sequencing. This plasmid contains an in-frame translational fusion of glutathione-S-transferase (GST), from pGEX-2T, and the *crhA* open reading frame. A thrombin cleavage

Figure 3.7. Flowsheet of site specific mutagenesis and isolation of the DNA fragment encoding the *crhA* open reading frame. Shown are the steps involved in creating a *Bam*HI site immediately preceeding the 5' end and an *Eco*RI site 100 bp downstream of the 3' end of the *crhA* open reading frame, as well as steps leading to the purification of this region of DNA. For details, see Section 2.4.1. The pWM75 plasmid shown is in a denatured, single stranded form. The dark area indicates the vector, pBluescript KS+ and the shaded area indicates the *crhA* open reading frame. Diagrams are not to scale.



Polymerization and Ligation

Transform mutagenesis reaction into *E. coli* DH5α

Inoculate cells into 5 ml LB media and grow overnight

Small-scale plasmid preparation

Digest with both *Bam*HI and *Eco*RI

Gel purify 1.4 kb band

site is encoded at the point of the fusion. The calculated molecular weight of the fusion protein was 74 kDa, composed of the 26 kDa GST and the 47 kDa CrhA peptides.

While all of the cloning steps were performed in *E. coli* DH5α, pGX29 was overexpressed in *E. coli* JM109 because *E. coli* DH5α overproduces a 70 kDa protein in response to abnormal protein synthesis which may mask the fusion protein.

Initial experiments focussed on optimizing overexpression. It was determined that using 2X YT-G media for growth did not efficiently repress expression under non-inducing conditions. For this reason, Trypticase® soy broth (TSB) containing 2% glucose was used for subsequent experiments. This medium is thought to contain absolutely no inducer whereas 2X YT-G may in fact contain traces of it. Growing the cells in TSB resulted in expression of the 74 kDa fusion protein only in the presence of inducer (Figure 3.8, lanes 1 and 2).

It was also determined that culture aeration was important for high levels of overexpression. Growing 150 mL cultures in 500 mL flasks did not result in significant overexpression, however, 20 mL cultures in 250 mL flasks grown with vigorous shaking resulted in very high levels of fusion protein production.

Once overexpression conditions were optimized, protein solubility was the next concern. In the protocol described by Pharmacia, TritonX-100 (1%) is used in order to solubilize the fusion protein. Using this method, cells were lysed first and then TritonX-100 added. Analysis of the supernatant and pellet after centrifugation showed that most, if not all of the fusion protein was in the pellet fraction, indicating insolubility. A second method for solubilization, described by Frangioni and Neel (1993) was then used. This procedure

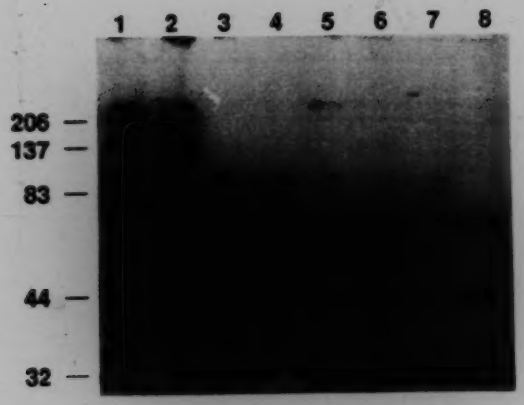


Figure 3.8. Overexpression of CrhA in *E. coli*. 10% Polyacrylamide gel showing various stages in the purification attempts of the GST/CrhA fusion protein: uninduced *E. coli* JM109 (pGX29) cells grown in TSB containing 2% glucose (lane 1); *E. coli* JM109 (pGX29) cells grown in TSB containing 2% glucose induced with 0.5 mM IPTG for 3 h (lane 2); pellet (lane 3) and supernatant (lane 4) of 1.5% sarkosyl-solubilized cell lysate after centrifugation; 2% sarkosyl-solubilized cell lysate cleared by centrifugation (lane 5); pellet (lane 6) and supernatant (lane 7) of cells lysed without sarkosyl and centrifuged at low speed (5000 rpm); and thrombin cleavage experiment (lane 8). For details see text. Molecular weight markers (Bio-Rad Kaleidoscope) are indicated in kDa.

is based on cell lysis in the presence of 1-2% sarkosyl to solubilize inclusion bodies. TritonX-100 is added in order to sequester the sarkosyl, which would otherwise interfere with subsequent affinity chromatography steps. Using 1.5% sarkosyl and 3% TritonX-100, the fusion protein was found to be equally divided 50:50 between the pellet and supernatant fractions (Figure 3.8, lanes 3 and 4). Because at least a portion of the fusion protein was solubilized using this method, it was used for preparation of the cell lysate. In some experiments, the amount of sarkosyl was increased to 2% in an effort to solubilize even more of the protein although the extent of the increased solubility was not determined. It was found however, that no real difference between lysate samples solubilized with 1.5% or 2% sarkosyl could be visually detected on gels (Figure 3.8, compare lanes 4 and 5). In addition, growing the cells at lower temperature (30°C and room temperature) did not result in any increase in solubility (data not shown).

Once the fusion protein was at least partially solubilized, attempts were made to purify it from the cell lysate by affinity chromatography using Glutathione Sepharose 4B beads. A batch binding method described by Pharmacia, in which the lysate and beads are incubated with gentle mixing, was used. The beads were then collected and the pre- and post-binding lysate compared to determine the extent of the binding. Incubation of the binding reaction mixture at room temperature for 30 min or at 4°C for 45 min to 4 h resulted in no binding (data not shown). There are several explanations for the lack of binding including: the use of excessive detergent (sarkosyl and/or TritonX-100), an insufficient use of TritonX-100 to sequester the sarkosyl, or a conformational alteration of the GST portion of the fusion protein.

The first two possibilities were addressed by a number of means. Excess detergent was removed by dialysis of the cell lysate against STE using dialysis tubing. A number of different TritonX-100 concentrations were next used in an attempt to sequester the sarkosyl; 2, 3, 4 and 5% TritonX-100 were used with 1.5% sarkosyl while 4, 5 and 6% TritonX-100 were used in the presence of 2% sarkosyl. Finally, the 2% sarkosyl-solubilized lysate with 4% TritonX-100 was incubated for 30 min at 4°C prior to adding the Glutathione Sepharose 4B. Despite these varied approaches, no significant binding of the fusion protein was observed.

It is possible that the sarkosyl/TritonX-100 combination was somehow interfering with binding. As a result, other detergents were tested. Pharmacia indicates that 0.03% SDS, 2% Tween and 1% CTAB are all compatible with binding. Of these, only 1% CTAB was able to solubilize the fusion protein, although to a lesser extent than sarkosyl (using visual estimation from protein gels approximately 25% of the fusion protein was soluble whereas sarkosyl facilitated 50% solubilization) (data not shown). Incubation of the 1% CTAB-solubilized lysate with Glutathione Sepharose 4B beads for 2 h at either room temperature or 4°C again resulted in no binding.

Two controls were done during the course of these binding experiments. GST alone, in the presence of 2% sarkosyl and with the addition of 3% TritonX-100, did not bind to Glutathione Sepharose 4B beads. In contrast, GST solubilized with 1% TritonX-100 did bind to the beads. This result indicated that the sarkosyl would have to be removed and not simply sequestered in order for fusion protein binding to be possible, barring any conformational problems with the fusion protein.

Lysate containing fusion protein solubilized with 1.5% sarkosyl was therefore dialyzed against STE using a Pierce Slide-A-Lyzer® Dialysis Cassette. The lysate collected after dialysis was very cloudy, suggesting the formation of a precipitate. Analysis of this precipitate showed that the fusion protein was present, suggesting that sarkosyl is required for solubility.

Since no binding of the fusion protein could be achieved, an alternative was the isolation of inclusion bodies, which should be essentially pure fusion protein. Induced cells were harvested, washed and resuspended in STE. The suspension was incubated with lysozyme, DTT added, and the suspension passed through a French Press. Low speed centrifugation (5000 rpm) concentrated all of the fusion protein into the pellet fraction along with a number of other proteins (Figure 3.8, lanes 6 and 7). In fact, the protein pattern in this fraction looked almost identical to the sarkosyl-solubilized cell lysate. In an attempt to remove any unlysed cells that may be responsible for the contaminating protein, the lysate was either passed through the French Press three times or passed through the French Press twice and sonicated. Neither modification resulted in a decrease in contaminating protein (data not shown).

The inclusion bodies were also washed with either STE or 0.25% sarkosyl in an attempt to remove contaminating proteins. While the STE wash had no effect, the 0.25% sarkosyl wash extracted all of the proteins into the soluble fraction. Since this low concentration of sarkosyl was able to solubilize the fusion protein, binding of this lysate, along with 2% TritonX-100, to the Glutathione Sepharose 4B beads was attempted. As previously observed, no binding occurred.

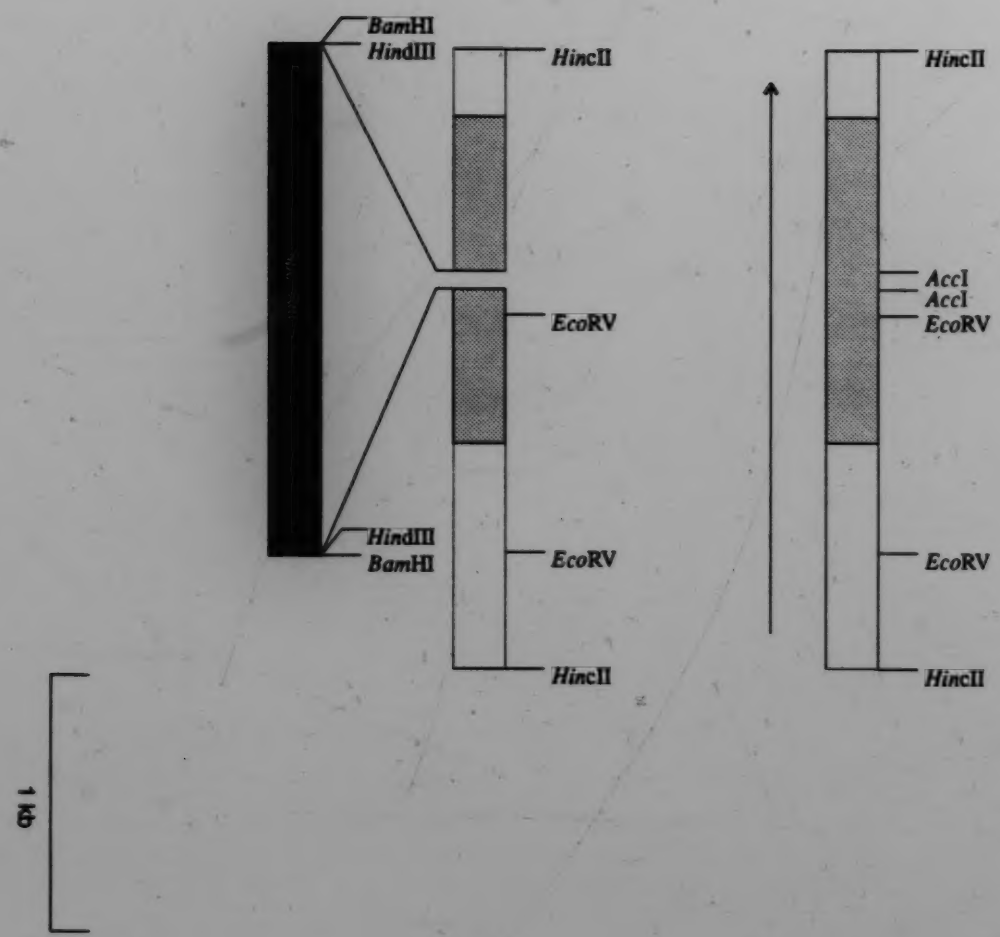
Although the fusion protein could not be purified from the cell lysate, a thrombin cleavage experiment was performed using 1.5% sarkosyl-solubilized lysate. Incubation of this lysate with thrombin at room temperature overnight resulted in the appearance of three new bands: one slightly smaller than the fusion protein, one slightly smaller than the expected size of full-length CrhA protein, and one corresponding to GST (Figure 3.8, lane 8). The GST (26 kDa) was run off and is not visible on the gel in Figure 3.8, but was seen on other gels run with the same reaction. These results suggest that there is an additional thrombin cleavage site within the fusion protein, somewhere within the 3' end of CrhA. Analysis of the deduced amino acid sequence indicated the presence of several putative thrombin cleavage sites within the CrhA open reading frame, however the precise site of cleavage was not determined.

3.5 INACTIVATION OF *crhA*

In order to study the physiological function of *crhA*, cyanobacterial *crhA*⁻ insertional inactivation mutants were generated. The resulting phenotypic effect was then monitored in an attempt to elucidate the role this gene performs in *A. variabilis* UTCC 387.

The *crhA* inactivation was first constructed *in vitro*, a schematic diagram of which is shown in Figure 3.9. The 2418 bp *Hin*cII/*Hin*cII fragment from pWM75 was cloned into *Hin*cII digested pBluescript KS+, creating pWM753 (Figure 3.9A). The *crhA* open reading frame on this plasmid was inactivated by the deletion of the 71 bp *Acc*I/*Acc*I fragment and insertion of the 2 kb spectinomycin/streptomycin resistance-encoding cassette from pHP45Ω. This plasmid, pKSINAC, thus contains a deletion of coding sequence and an insertion of

Figure 3.9. Construction of the *crhA* inactivation plasmid. A: Restriction map of the 2418 bp *Hin*cII fragment comprising pWM753. The shaded area indicates the *crhA* open reading frame. The arrow underneath the map indicates the direction of transcription and is not to scale. **B:** Physical map of pKSINAC showing the deletion of the *Acc*I/*Acc*I fragment and the insertion of the spectinomycin/streptomycin resistance (Spc^R/Sm^R) cassette (black rectangle). The scale is shown for 1 kb.


B
A
BamHI
HindIII
HincII
EcoRV
HindIII
BamHI
EcoRV
HincII
HincII
AccI
AccI
EcoRV
EcoRV
HincII
1kb

foreign DNA within the *crhA* open reading frame (Figure 3.9B). The *crhA* inactivation construct was removed from pBluescript KS+ and transferred to pBR322, creating pBRINAC. This was done because pBR322 contains a *bom* (*b*asis *o*f *m*obility) site, required for cyanobacterial conjugation, whereas pBluescript KS+ does not. The insert of pBRINAC is identical to pKSINAC (Figure 3.9B), except that the former has lost the *Hin*cII sites at either end. Both pKSINAC and pBRINAC were confirmed by restriction analysis.

In order to inactivate *crhA in vivo*, pBRINAC first had to be transferred to *A. variabilis* UTCC 387, at which time homologous recombination could occur between the inactivated copy of *crhA* on the plasmid and the wild type copy in the genome. pBR322 was introduced into *A. variabilis* UTCC 387 by conjugation. This plasmid was first transformed into *E. coli* HB101(pRL623); pRL623 is a helper plasmid that encodes the *mob* protein as well as methylases to protect *Ava*I, *Ava*II, and *Ava*III sites. The conjugal strain used was *E. coli* JE2571 (RP4), which provided the proteins necessary for transfer. The two *E. coli* strains and the wild type *A. variabilis* UTCC 387 were mixed and incubated at 30°C without shaking under continuous light overnight. The next day, transconjugants were selected on BG-11 plates containing spectinomycin and streptomycin. After 12 days incubation, a number of green smeared areas, where single colonies could not be distinguished, appeared on the plates. Restreaking portions of eight of these smeared areas resulted in growth of individual colonies for all but one. In addition, a number of individual colonies arose on the original plates, 19 days after plating.

Sixty-four (64) individual colonies from both the re-streaked smears and 19 day growth on the original plates were chosen for further study. These colonies were subcultured

several times in an attempt to segregate inactivated and wild type genomes. In order to distinguish single crossover from double crossover events, these colonies were plated onto media containing ampicillin. Resistance to this antibiotic would be indicative of a single crossover event as the vector pBR322 encodes β-lactamase. The majority of the colonies were ampicillin resistant. Eight of the colonies showed very poor growth on the ampicillin plates, suggesting they may be ampicillin sensitive. All eight of these clones were chosen as potential double crossover mutants.

These putative ampicillin resistant clones all grew in the presence of spectinomycin and streptomycin on both BG-11 plates and in liquid culture. The rate of growth of mutants of solid BG-11 media was comparible with that observed for wild type cells, suggesting no obvious growth defect. In contrast, microscopic analysis of liquid cultures showed that the mutant filaments were truncated, having lengths of 1-8 cells (Figure 3.10B), while wild type cells have filaments of 40-100 cells (Figure 3.10A).

In order to determine if the eight mutants were in fact a result of double crossover events, Southern analysis was performed. Genomic DNA from each mutant was isolated and digested with *Hin*cII. Wild type genomic DNA, also digested with *Hin*cII, served as a control. DNA fragments were separated by electrophoresis, transferred to a nylon membrane and probed with the 936 bp *Eco*RV/*Eco*RV insert from pWM752 (Figure 3.3B). The resulting autoradiogram is shown in Figure 3.11. A single hybridizing band of approximately 2.4 kb is observed in wild type DNA (Figure 3.11, lane 1), as expected (Figure 3.9A). All of the mutants (Figure 3.11, lanes 2-9) show identical results. If a double crossover event had occurred and all genomes in each cell were inactivated, only a single

Figure 3.10. Morphology of wild type (A) and mutant (B) *A. variabilis* UTCC 387 cells.

Photographs were taken microscopically under 400X power.

A

B

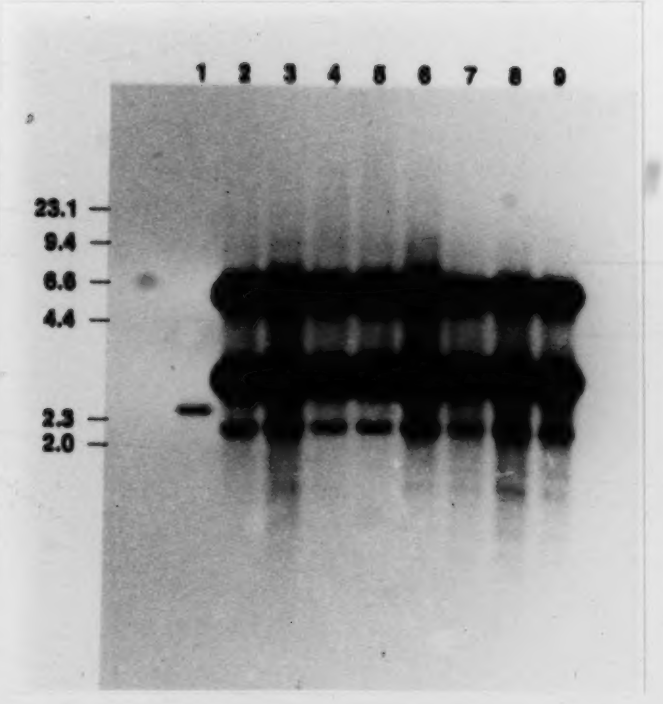
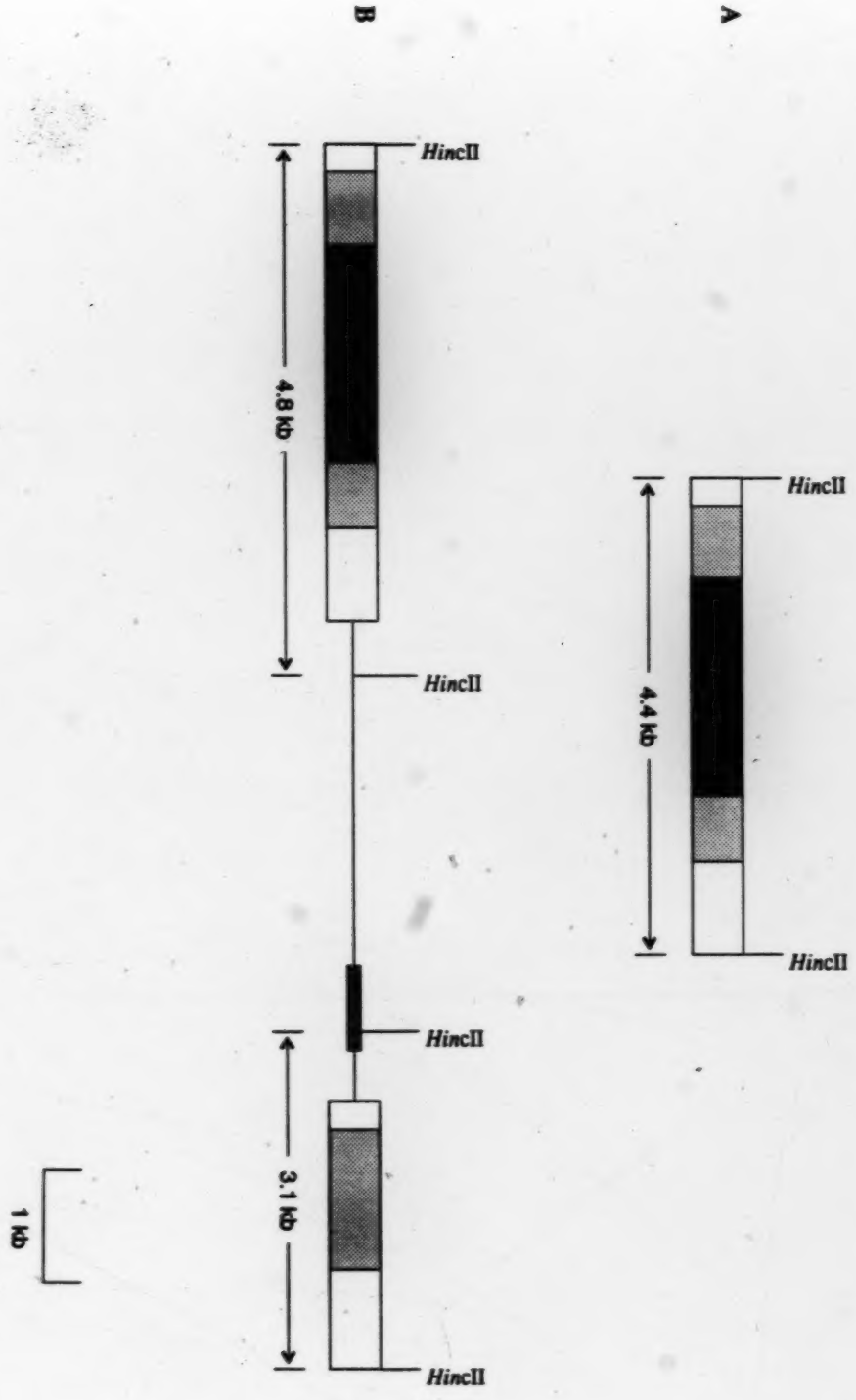


Figure 3.11. Southern analysis of spectinomycin/streptomycin resistant *A. variabilis* UTCC 387 genomic DNA. Genomic DNA (2 μg) from wild type (lane 1) and eight spectinomycin/streptomycin resistant clones (lanes 2-9) was digested with *Hin*cII, separated by electrophoresis and subjected to Southern transfer. The membrane was probed with the 936 bp *Eco*RV/*Eco*RV insert from pWM752 (Figure 3.3B). Size markers (*Hin*dIII digested lambda DNA) are indicated in kb.

hybridizing band of approximately 4.4 kb (2.4 kb from pWM753 plus 2.0 kb from the spectinomycin/streptomycin resistance cassette) would be seen (Figure 3.12A). In contrast, these mutants show two strongly hybridizing bands (4.8 kb and 3.1 kb), the sizes of which are consistent with a single crossover event (Figure 3.12B). The identity of the lower, more weakly hybridizing band (approximately 2.2 kb) in each of these lanes is unknown. Thus, it appears that the eight inactivation mutants are not null mutants, as they contain both a wild type and an inactivated copy of the *crhA* gene. Interestingly, they still exhibit an altered growth phenotype in liquid culture.

Figure 3.12. *In vivo* inactivation of *crhA*. Restriction maps of the genome of *A. variabilis* UTCC 387 in the area of the *crhA* gene showing two possible modes of recombination: double crossover event **(A)** and single crossover event **(B)**. Only *Hin*cII sites are shown. Sizes are indicated for hybridizing fragments. The shaded areas in both diagrams indicate the *crhA* open reading frame. The single line in **(B)** indicates the vector pBR322 and the thick bar within this sequence represents the ampicillin resistance gene. The black rectangles show the site of the spectinomycin/streptomycin resistance cassette for each case. The scale is shown for 1 kb.


B
A
HincII
4.8 kb
HincII
HincII
3.1 kb
HincII
1 kb
HincII
4.4 kb
HincII

CHAPTER FOUR
DISCUSSION

A fragment of *A. variabilis* UTCC 387 genomic DNA was isolated by homologous probing with the fragment gwo29. gwo29 was synthesized in a PCR reaction which utilized total genomic DNA as a template and degenerate primers corresponding to two of the highly conserved amino acid motifs found in DEAD box RNA helicases (Magee, 1995). During screening of *A. variabilis* UTCC 387 *Xba*I genomic clone banks, three positive clones were identified. On the basis of *Xba*I restriction analysis, two of these clones were shown to be identical with insert sizes of 4 kb. The third clone differed in that its insert size was approximately 4.4 kb. The isolation of clones with two different insert sizes was unexpected since a genomic Southern probed with the same PCR fragment revealed only a single hybridizing *Xba*I band (Magee, 1995). This signal was broad however, and since the two insert sizes are similar, it is possible that both bands were in fact present but due to strong hybridization, they appeared as one band. It is important to note that the DNA bands in the clone bank gel were better resolved than the bands in the genomic gel.

Growth of the 4.4 kb clone in liquid culture resulted in the eventual loss of the insert. The reason for the loss of this clone is unknown, as is its identity. Two explanations are possible. Firstly, the single genomic *Xba*I band may be a combination of two signals such that the 4.4 kb fragment is entirely unrelated to the 4 kb fragment. The former fragment may then encode a second, distinct RNA helicase that shows homology to gwo29. In this scenario, lethality to *E. coli* could be a result of the helicase itself, or to sequences surrounding the helicase. Secondly, the 4.4 kb fragment may in fact contain the 4 kb

fragment, but due to incomplete *Xba*I digestion, an additional 400 bp of DNA is present. Isolation and characterization of the 4.4 kb clone will be required to determine which of these possibilities is correct. All recovery attempts, including re-screening of the clone bank in *E. coli* DH5α, failed. However, the clone is still present in the clone bank DNA preparation used to transform *E. coli* DH5α, so it has not been lost entirely. Alternative methods for isolation may have to be employed such as propagation in a host other than *E. coli*, or construction of a new clone bank in a low copy number vector, if lethality is a result of high plasmid copy number.

As the two remaining clones were deemed to be the same on the basis of their identical insert sizes, only one was chosen for further study and designated pWM75. This plasmid was first characterized by restriction endonuclease mapping. Out of 19 restriction endonucleases used, only *Acc*I, *Eco*RV and *Hinc*II cut within the insert.

Mapping of the *Acc*I, *Eco*RV and *Hinc*II sites provided the first step in subcloning the 4 kb fragment into smaller pieces for sequencing. The constructed subclones were used to generate 1964 bp of sequence data located in the middle of the 4 kb fragment. Within this sequence an open reading frame of 1278 bp was found. The open reading frame has a coding capacity of 425 amino acids. Comparison of the deduced amino acid sequence to that determined for the PCR fragment showed that the two are not identical. Two factors could account for this difference. Firstly, the PCR fragment may have been derived from the same piece of genomic DNA that is present in pWM75, but that errors were introduced during PCR amplification. This is very plausible, since *Taq* DNA polymerase was used for the PCR, and lacks 3' to 5' exonuclease proofreading activity. Secondly, the differences may in

fact be real, in which case an additional RNA helicase, highly homologous to the one isolated and containing an identical sequence to the PCR product, exists in the *A. variabilis* UTCC 387 genome. This latter explanation could account for the 4.4 kb insert clone that was identified during clone bank screening. Again, isolation and characterization of this clone will be required to obtain a more definitive answer.

Analysis of the 425 amino acid sequence revealed that all nine of the diagnostic conserved RNA helicase amino acid motifs are present. As such, it was named *crhA*, for *c*yanobacterial *R*NA *h*elicase *A*. This designation is not intended to be permanent; a more specific and descriptive name can be assigned once it is characterized further and a cellular function is known.

Interestingly, in the *chrA* deduced amino acid sequence the SAT motif has been replaced by a FAT motif. This is the first description of FAT in the literature. The change is not a result of a sequencing error as both strands of DNA were (separately) sequenced at least twice; the PCR fragment also contained a FAT motif. This amino acid replacement could have resulted from a single nucleotide substitution. The phenylalanine codon in the *chrA* sequence is TTC. One of the six serine codons is TCC. Therefore, a transition from C to T could have occurred in the second position of this codon. In addition, the SAT to FAT change is not necessarily a characteristic of *A. variabilis* UTCC 387 RNA helicases, since PCR analysis yielded an additional, unrelated fragment, gwo4, that contains a SAT motif (Magee, 1995).

The replacement of SAT with FAT is unusual in that all known members of the DEAD box RNA helicase family contain the SAT motif, an indication that CrhA may

perform a specific function in cyanobacteria. In fact, DEXH family members also contain an altered SAT box, a TAT box (Pause and Sonenberg, 1993). Proteins in this latter group include those that are involved in genome replication in RNA viruses (Fuller-Pace, 1994), like the plum pox potyvirus C1 protein (Férnandez, *et al.*, 1995). A serine to threonine change, like that seen in DEXH proteins compared to DEAD box proteins is conservative in that both amino acids have uncharged polar side chains. In contrast, a serine to phenylalanine change, like that seen in CrhA, is nonconservative because an uncharged polar amino acid is replaced with a nonpolar amino acid. The significance of the change is unknown and needs to be investigated. Extrapolating from the data obtained by Pause and Sonenberg (1992) in their *in vitro* mutational analysis of purified recombinant eIF4A, the FAT motif may be responsible for unwinding a specific RNA duplex that would define this RNA helicase's specific cellular function. These authors examined the SAT box by replacing both the serine and threonine with alanine, creating an AAA motif. This mutation did not negatively affect ATP binding nor ATPase activities, but did abolish RNA unwinding (Pause and Sonenberg, 1992). It would be of interest to determine the effect of a serine to phenylalanine change on enzymatic activity in a well-studied RNA helicase like eIF4A. Such an analysis may lead to a greater insight into the purpose of this change.

A putative promoter region for the 1278 bp open reading frame was also present in the 1964 bp sequence. Comparison to other cyanobacteria promoters was difficult in that no specific cyanobacterial consensus sequences have been identified (Houmard, 1994). Houmard (1994) suggested that this is not due to a lack of consensus sequences, but rather because too few cyanobacterial promoters have been mapped. Many of the promoters that

have been studied are controlled by varying environmental and/or developmental factors, therefore their transcription is regulated quite differently than housekeeping promoters. Even so, Curtis and Martin (1994) state that most cyanobacterial promoters contain a conserved element at -10 from the transcription initiation site that conforms to the *E. coli* -10 consensus sequence, TATAAT; a number of these were identified in the 5' untranslated region of *crhA*. In contrast, no *E. coli* -35 consensus sequence (TTGACA) was found, consistent with the finding that less than half of the characterized cyanobacterial promoters contain this sequence and that even when it is present, its conservation is relatively weak (Curtis and Martin, 1994). Promoter mapping experiments will have to be performed in order to characterize the *chrA* promoter more extensively. A possible *E. coli* Shine-Dalgarno sequence was also identified 5' to the initiator codon, although no information could be found to relate this sequence to others found in cyanobacteria.

Upon comparing the amino acid sequence to other sequences found in the GenBank database, a number of known or putative RNA helicases were identified that showed significant identity. It had been hoped that the homology search would provide a clue to the function of this gene, perhaps by showing very significant similarity and/ or identity to an RNA helicase involved in a specific cellular pathway. The most closely related protein however, was RhlE from *E. coli*, suggesting that ChrA is the RhlE homologue in *A. variabilis* UTCC 387. Unfortunately, the function of RhlE is unknown (Ohmori, 1994). Only its nucleotide sequence and a null mutant, which showed no alteration in growth, have been described (Ohmori, 1994). The fact that the entire *chrA* deduced amino acid sequence, and not simply the conserved motifs, exhibits significant identity to a number of RNA

helicases provides even stronger evidence that this gene is, in fact, an RNA helicase.

While *crhA* has been tentatively identified as an RNA helicase on the basis of amino acid sequence homology, verification requires demonstration of ATP-dependent RNA unwinding and RNA-dependent ATPase activities. To that end, attempts were made to purify the CrhA protein for use in enzymatic assays. The Glutathione-S-Transferase (GST) Gene Fusion System from Pharmacia was used for this purpose. One of the first problems associated with this system was that the fusion protein was insoluble in cell lysates. Application of the procedure described by Frangioni and Neel (1993), in which a combination of sarkosyl and TritonX-100 are used as solubilization agents, resulted in approximately 50% solubility of the fusion protein. Unfortunately, none of the conditions tested, including varying the detergent concentrations, yielded successful affinity purification with Glutathione Sepharose 4B beads. A control in which native GST, solubilized by TritonX-100 only, was shown to bind to the affinity matrix eliminated the possibility that the matrix was the cause of the problem and instead, suggested altered GST conformation within the fusion protein. In a second control experiment, GST was solubilized by both sarkosyl and TritonX-100 but still did not bind to the matrix. This result suggested instead that binding was being prevented by the sarkosyl/TritonX-100 combination. Indeed, Frangioni and Neel (1993) found that sarkosyl did prevent binding to the affinity matrix, but that TritonX-100 could be added in order to sequester the sarkosyl and allow binding. The authors stated that an appropriate sarkosyl:TritonX-100 molar ratio must be empirically determined for each fusion protein. As the second control experiment was performed only once, it is possible that the correct ratio had not been used and that by determining this ratio,

binding could have been demonstrated. If this explanation is correct, then it once again points to altered GST conformation as the cause of no affinity matrix binding. Further support for this theory comes from the fact that when 1% CTAB is used to solubilize the fusion protein, binding still did not occur, even though the GST Gene Fusion System Manual (Pharmacia, 1993) states that binding has been achieved in the presence of this concentration of the detergent.

An alternative method for the purification of the fusion protein exploited its insoluble nature in the absence of sarkosyl. It had been hypothesized that lysing the cells in the absence of sarkosyl, collection of the insoluble fraction by low speed centrifugation, followed by extraction with sarkosyl would yield essentially pure protein, similar to that seen with *E. coli*-expressed actin (Frankel, *et al.*, 1991; McNally, *et al.*, 1991). Unfortunately, the insoluble fraction, which was thought to consist of inclusion bodies of pure fusion protein, was in fact contaminated with a large number of other proteins, all of which were sarkosyl soluble. Despite numerous attempts, the protein could not be purified any further.

Thrombin cleavage experiments showed that in addition to the engineered cleavage site between GST and CrhA, an additional site must be present somewhere within the carboxy- terminus of CrhA. A slightly smaller, presumably truncated, version of the fusion protein and of the CrhA cleavage product was seen on polyacrylamide gels run with the thrombin reactions. As a GST band of expected size was also seen, cleavage could not have occurred in the amino terminus. A number of potential cleavage sites in the amino acid sequence of CrhA were identified by sequence analysis (Pharmacia, 1993), but the exact one that was being cleaved was not identified.

Because purification of CrhA using the GST Gene Fusion System was unsuccessful, attempts are currently in progress to clone the open reading frame into a different expression vector, pQE-9 from QIAGEN. This vector creates a translational fusion of six histidine residues with the protein of interest. The fusion can then be purified by affinity chromatography on a nickel column. One advantage of this system, in contrast to the pGEX system is that unlike GST, the histidine tag is not immunogenic and need not be removed prior to making antibodies if the protein were being purified for this purpose. More importantly, the interaction between the tag and the nickel column is not based on tertiary structure so the fusion protein can be purified in the presence of strong denaturants and detergents. Since the most likely explanation for the lack of Glutathione Sepharose bead binding is thought to be an altered GST conformation in the GST/CrhA fusion, the histidine tag is expected to yield greater success.

In order to study the physiological activity of *crhA*, the gene was first inactivated *in vitro* by deletion of a small internal portion of the open reading frame and insertion of a spectinomycin resistance cassette. *In vivo* inactivation was achieved by homologous recombination between the inactivated construct and the wild-type genomic copy. Southern analysis of eight inactivants, each of which showed very poor growth on plates containing ampicillin, revealed that they were all a result of single crossover events that integrated the entire plasmid, not double crossovers. Therefore, each contained both a wild type and an inactivated copy of the *crhA* gene. The reason for poor growth on ampicillin-containing plates is unknown. The remaining 56 exconjugants grew well on plates with ampicillin; growth was indistinguishable from that on plates containing spectinomycin and

streptomycin. Presumably these latter exconjugants also arose as a result of single crossover events. It would be of interest to do a Southern analysis of these ampicillin resistant colonies in order to compare results for the two colony types. Perhaps a difference would be seen that could account for the differences in growth on ampicillin containing plates.

The failure to obtain any double recombinants is not altogether surprising. Cai and Wolk (1990) stated that when homologous DNA is transferred to *Anabaena* sp. by conjugation, single crossover events occur at a much higher frequency than double crossover events. The authors also stated that exhaustive screens have failed to isolate double recombinants in *Anabaena* sp. strain PCC 7120 when the size of the homologous region was below 4 kb (Cai and Wolk, 1990). This latter point may have been a key factor in the failure to isolate double crossover recombinants in this case, as the size of the homologous region was only 2347 bp, with the spectinomycin resistance cassette inserted near the centre. In order to completely inactivate the gene, it will be necessary to repeat the conjugation experiment using conditions that increase recovery of double crossover events either by increasing the size of the homologous region or by selecting specifically for double recombinants, perhaps by using the conditionally lethal *sacB* gene in the vector and growing exconjugants on plates containing 5% sucrose (Cai and Wolk, 1990).

Although complete inactivation was not achieved, a phenotypic change was evident, suggesting a dominant mutation had occurred. Specifically, the growth morphology in liquid culture was altered. Whereas wild type *A. variabilis* UTCC 387 grown in liquid displays long filaments of 40-100 cells, the eight inactivants showed very short filaments (less than 10 cells) and were often unicellular. This result suggests a role for *crhA* in filamentation,

possibly in cell wall or cell membrane metabolism. The dominant nature of the mutation could be explained if the inactivated *crhA* gene product is interfering with the wild type protein by sequestering its substrate.

Interestingly, preliminary Northern analyses have indicated that this gene is expressed only under cold growth conditions (22°C) (Dr. D. Chamot, unpublished results). Cold-regulated growth has not been extensively studied in *Anabaena* sp.. Sato has described a low temperature-induced gene, *lti2* from *Anabaena variabilis* M3 that shows homology to α-amylases (1992), as well as a family of cold-regulated RNA-binding protein genes in the same strain (1994; 1995). The latter genes encode small proteins (93 to 107 amino acid residues) that contain RNA-recognition motifs (Sato, 1995). By far the most work on low-temperature effects on *Anabaena* sp. has focussed on changes in the cell membrane. Sato, *et al.* (1979), determined that although the lipid composition of *A. variabilis* M3 was unaffected by a 38°C to 22°C change in growth temperature, the fatty acid composition, specifically the degree of saturation, was affected. The authors found that the 18-carbon fatty acids in particular increased their unsaturation in response to the temperature drop, although the 16-carbon fatty acids were affected to a certain extent as well. These findings were confirmed by Sato and Murata (1980), who concluded that the change in fatty acid composition was a mechanism for the regulation of membrane fluidity. Indeed, the phase transition temperature of phospholipid bilayers is inversely proportional to the degree of unsaturation (Sato, *et al.*, 1979). Later, Sato and Murata (1981) demonstrated that fatty acid desaturation in response to a temperature drop was not dependent on *de novo* fatty acid synthesis. In contrast, they found that RNA and protein synthesis were required, suggesting

that desaturases were synthesized as a result of the temperature shift. Although the desaturase genes have not yet been cloned from this organism, at least one, *desA*, has been identified in *Synechocystis* sp. strain PCC 6803 (Wada, *et al.*, 1990). Expression of *desA* is increased with a decrease in temperature, initially suggesting low temperature induction (Murata and Wada, 1995). Further studies using catalytic hydrogenation of unsaturated fatty acids in membrane lipids revealed that *desA* is similarly induced, suggesting that transcription is regulated in response to membrane fluidity and not to the growth temperature *per se* (Murata and Wada, 1995). The signalling mechanism that responds to the physical properties of the membrane has not yet been elucidated. Since the *crhA* inactivation mutants showed altered growth morphology and the *crhA* gene appears to be induced under cold growth conditions, it is possible that this RNA helicase plays a role in cell membrane synthesis and/or in regulating changes in membrane fluidity, perhaps by interacting with these desaturases.

A role for an RNA helicase in the cold shock response is not unprecedented. Jones, *et al.* (1996) found that CsdA (formerly DeaD) is induced when an *E. coli* culture is shifted from 37°C to 15°C. A *csdA* deletion mutant slowed cell growth and impaired the synthesis of a number of proteins at low temperature (Jones, *et al.*, 1996). The authors hypothesized that CsdA is required to increase translational efficiencies by unwinding stable secondary structures that form at low temperature. In addition, the *csdA* deletion strain formed long filamentous cells after 96 hours of growth at 15°C (Jones, *et al.*, 1996). This finding is in contrast to the case seen in *crhA*, in that in the former, the mutation of the RNA helicase results in filamentation whereas in the latter, the inactivation of *crhA* results in a loss of

filamentation. It will be of interest to see what effect the *chrA* mutants have on the growth of *A. variabilis* UTCC 387 at low temperature.

Obviously much more work is needed to determine the exact role *crhA* plays in *A. variabilis* UTCC 387. The fact that it may be involved in a cold-shock or cold-tolerance response though, raises exciting possibilities. CrhA may indeed perform a very specific function related to the ability of *A. variabilis* UTCC 387 to respond to low temperature. If proven, this opens up a wide spectrum of experiments designed to identify specific RNA targets for CrhA. These targets could include transcripts of known cold shock genes like *lti2* and the cold-regulated RNA-binding proteins. Identification of target RNAs for CrhA would allow their interaction to be studied in more detail, at both molecular and biochemical levels. In addition, this organism may be used as a model system to study cold adaptation in higher plants. Indeed, the biochemical mechanisms of desaturation of 16-carbon fatty acids have been shown to be similar in the two groups of organisms (Sato and Murata, 1982). Whatever the exact function CrhA performs, it indicates that RNA helicases are not necessarily involved in general cellular functions like that seen in the prototype eIF4A's role in eukaryotic translation initiation in which any mRNA species is a substrate.

BIBLIOGRAPHY

Abramson, R.D., T.E. Dever, T.G. Lawson, B.K. Ray, R.E. Thach, and W.C. Merrick. 1987. The ATP-dependent interaction of eukaryotic initiation factors with mRNA. J. Biol. Chem. **262:**3826-3832.

Allen, M.M. 1968. Simple conditions for growth of unicellular blue-green algae on plates. J. Phycol. **4:**1-4.

Bancroft, I., C.P. Wolk, and E.V. Oren. 1989. Physical and genetic maps of the genome of the heterocyst-forming cyanobacterium *Anabaena* sp. strain PCC 7120. J. Bacteriol. **171:**5940-5948.

Bauer, C.C. and R. Haselkorn. 1995. Vectors for determining the differential expression of genes in heterocysts and vegetative cells of *Anabaena* sp. strain PCC 7120. J. Bacteriol. **177:**3332-3336.

Böddeker, N., K. Stade, and F. Franceschi. 1997. Characterization of DbpA, an *Escherichia coli* DEAD box protein with ATP independent RNA unwinding activity. Nucleic Acids Res. **25:**537-544.

Bolivar, F., R.L. Rodriguez, R.J. Greene, M.C. Betlach, H.L. Heyneker, and H.W. Boyer. 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene. **2:**95-113.

Buikema, W.J. and R. Haselkorn. 1993. Molecular genetics of cyanobacterial development. Annu. Rev. Plant Physiol. Plant Mol. Biol. **44:**33-52.

Cai, Y. and C.P. Wolk. 1990. Use of a conditionally lethal gene in *Anabaena* sp. strain PCC 7120 to select for double recombinants and to entrap insertion sequences. J.

Bacteriol. **172:**3138-3145.

Cai, Y. and C.P. Wolk. 1997a. Nitrogen deprivation of *Anabaena* sp. strain PCC 7120 elicits rapid activation of a gene cluster that is essential for uptake and utilization of nitrate. J. Bacteriol. **179:**258-266.

Cai, Y. and C.P. Wolk. 1997b. *Anabaena* sp. strain PCC 7120 responds to nitrogen deprivation with a cascade-like sequence of transcriptional activations. J. Bacteriol. **179:**267-271.

Carrasco, C.D. and J.W. Golden. 1995. Two heterocyst-specific DNA rearrangements of *nif* operons in *Anabeana cylindrica* and *Nostoc* sp. strain Mac. Microbiol. **141:**2479-2487.

Carrasco, C.D., K.S. Ramaswamy, T.S. Ramasubramanian, and J.W. Golden. 1994. *Anabaena xisF* gene encodes a developmentally regulated site-specific recombinase. **Genes Dev. 8:74-83.**

Curtis, S.E. and J.A. Martin. 1994. The transcription apparatus and the regulation of transcription termination, p. 613-639. *In* D.A. Bryant (ed.), The molecular biology of cyanobacteria. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Datta, N., R.W. Hedges, E.J. Shaw, R.B. Sykes, and M.H. Richmond. 1971. Properties of an R factor from *Pseudomonas aeruginosa*. J. Bacteriol. **108:**1244-1249.

Elhai, J., A. Vepritskiy, A.M. Muro-Pastor, E. Flores, and C.P. Wolk. 1997. Reduction of conjugal transfer efficiency by three restriction activities of *Anabaena* sp. strain PCC 7120. J. Bacteriol. **179:**1998-2005.

Elhai, J. and C.P. Wolk. 1988. Conjugal transfer of DNA to cyanobacteria. Methods

Enzymol. **167:**747-754.

Elhai, J. and C.P. Wolk. 1990. Developmental regulation and spatial pattern of expression of the structural genes for nitrogenase in the cyanobacterium *Anabaena*. EMBO J. **9:**3379-3388.

Fay, P. 1983. The Blue-Greens. Edward Arnold, Baltimore, MD.

Feinberg, A.P. and B. Vogelstein. 1983. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132:**6-13.

Férnandez, A., S. Laín, and J.A. García. 1995. RNA helicase activity of the plum pox potyvirus C1 protein expressed in *Escherichia coli*. Mapping of an RNA binding domain. Nucleic Acids Res. **23:**1327-1332.

Frangioni, J.V. and B.G. Neel. 1993. Solubilization and purification of enzymatically active glutathione-S-transferase (pGEX) fusion proteins. Anal. Biochem. **210:**179-187.

Frankel, S., R. Sohn, and L. Leinwand. 1991. The use of sarkosyl in generating soluble protein after bacterial expression. Proc. Natl. Acad. Sci. USA. **88:**1192-1196.

Frost, L.S. and W. Paranchych. 1988. DNA sequence analysis of point mutations in *traA*, the F pilin gene, reveal two domains involved in F-specific bacteriophage attachment. Mol. Gen. Genet. **213:**134-139.

Fuller-Pace, F.V. 1994. RNA helicases: modulators of RNA structure. Trends Cell Biol. **4:**271-274.

Fuller-Pace, F.V., S.M. Nicol, A.D. Reid, and D.P. Lane. 1993. DbpA: a DEAD box protein specifically activated by 23S rRNA. EMBO J. **12:**3619-3626.

Golden, J.W., C.D. Carrasco, M.E. Mulligan, G.J. Schneider, and R. Haselkorn. 1988. Deletion of a 55-kilobase-pair DNA element from the chromosome during heterocyst differentiation of *Anabaena* sp. strain PCC 7120. J. Bacteriol. **170:**5034-5041.

Golden, J.W., M.E. Mulligan, and R. Haselkorn. 1987. Different recombination site specificity of two developmentally regulated genome rearrangements. Nature **327:**526-529.

Golden, J.W. and D.R. Wiest. 1988. Genome rearrangement and nitrogen fixation in *Anabaena* blocked by inactivation of *xisA* gene. Science **242:**1421-1423.

Golden, S.S. 1988. Mutagenesis of cyanobacteria by classical and gene-transfer-based methods. Methods Enzymol. **167:**714-727.

Gorbalenya, A.E. and E.V. Koonin. 1993. Helicases: amino acid sequence comparisons and structure-function relationships. Curr. Opin. Struct. Biol. **3:**419-429.

Gorbalenya, A.E., E.V. Koonin, A.P. Donchenko, and V.M. Blinov. 1989. Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. Nucleic Acids Res. **17:**4713-4730.

Gray, M.W. and W.F. Doolittle. 1982. Has the endosymbiont hypothesis been proven? Microbiol. Rev. **46:**1-42.

Grifo, J.A., R.D. Abramson, C.A. Satler, and W.C. Merrick. 1984. RNA-stimulated ATPase activity of eukaryotic initiation factors. J. Biol. Chem. **259:**8648-8654.

Gulyas, K.D. and T.F. Donahue. 1992. *SSL2*, a suppressor of a stem-loop mutation in the *HIS4* leader encodes the yeast homolog of human *ERCC-3*. Cell. **69:**1031-1042.

Gururajan, R., H. Perry-O'Keefe, D.A. Melton, and D.L. Weeks. 1991. The *Xenopus*

localized messenger RNA An3 may encode an ATP-dependent RNA helicase. Nature. **349:**717-719.

Haselkorn, R. 1991. Genetic systems in cyanobacteria. Methods Enzymol. **204:**418-430.

Haselkorn, R. 1992. Developmentally regulated gene rearrangements in prokaryotes. Annu. Rev. Genet. **26:**113-130.

Haselkorn, R., J.W. Golden, P.J. Lammers, and M.E. Mulligan. 1986. Developmental rearrangement of cyanobacterial nitrogen-fixation genes. Trends Genet. **2:**255-259.

Herdman, M., M. Janvier, R. Rippka, and R.Y. Stanier. 1979. Genome size of cyanobacteria. J. Gen. Microbiol. **111:**73-85.

Herdman, M., M. Janvier, J.B. Waterbury, R. Rippka, R.Y. Stanier, and M. Mandel. 1979. Deoxyribonucleic acid base composition of cyanobacteria. J. Gen. Microbiol. **111:**63-71.

Hodgman, T.C. 1988. A new superfamily of replicative proteins. Nature. **333:**22-23.

Houmard, J. 1994. Gene transcription in filamentous cyanobacteria. Microbiol. **140:**433-441.

Iggo, R., S. Picksley, J. Southgate, J. McPheat, and D.P. Lane. 1990. Identification of a putative RNA helicase in *E. coli*. Nucleic Acids Res. **18:**5413-5417.

Jackman, D.M. and M.E. Mulligan. 1995. Characterization of a nitrogen-fixation (*nif*) gene cluster from *Anabaena azollae* 1a shows that closely related cyanobacteria have highly variable but structured intergenic regions. Microbiol. **141:**2235-2244.

Jones, P.G., M. Mitta, Y. Kim, W. Jiang, and M. Inouye. 1996. Cold shock induces a major ribosomal-associated protein that unwinds double-stranded RNA in

Escherichia coli. Proc. Natl. Acad. Sci. USA **93:**76-80.

Kalman, M., H. Murphy, and M. Cashel. 1991. *rhlB*, a new *Escherichia coli* K-12 gene with an RNA helicase-like protein sequence motif, one of at least five such possible genes in a prokaryote. New Biologist **3:**886-895.

Kaneko, T., A. Tanaka, S.Sato, H. Kotani, T. Sazuka, N. Miyajima, M. Sugiura, and S. Tabata. 1995. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome. DNA Res. **2:**153-166.

Lammers, P.J., J.W. Golden, and R. Haselkorn. 1986. Identification and sequence of a gene required for a developmentally regulated DNA excision in *Anabaena*. Cell **44:**905-911.

Lasko, P.F. and M. Ashburner. 1988. The product of the *Drosophila* gene *vasa* is very similar to eukaryotic initiation factor-4A. Nature. **335:**611-617.

Leroy, P., P. Alzari, D. Sassoon, D. Wolgemuth, and M. Fellous. 1989. The protein encoded by a murine male germ cell-specific transcript is a putative ATP-dependent RNA helicase. Cell. **57:**549-559.

Linder, P., P.F. Lasko, P. Leroy, P.J. Nielsen, K. Nishi, J. Schnier, and P.P. Slonimski. 1989. Birth of the D-E-A-D box. Nature **337:**121-122.

Magee, W. 1995. RNA helicase gene cloning and molecular analysis. Undergraduate Thesis Project Report. University of Alberta. p. 21.

Marmur, J. 1961. A procedure for the isolation of deoxyribonucleic acid from micro-organisms. J. Mol. Biol. **3:**208-218.

McNally, E., R. Sohn, S. Frankel, and L. Leinwand. 1991. Expression of myosin and actin in *Escherichia coli*. Methods Enzymol. **196:**368-389.

Méthot, N., A. Pause, J.W.B. Hershey, and N. Sonenberg. 1994. The translation initiation factor eIF-4B contains an RNA binding region that is distinct and independent from its ribonucleoprotein consensus sequence. Mol. Cell Biol. **14:**2307-2316.

Miczak, A., V.R. Kaberdin, C. Wei, and S. Lin-Chao. 1996. Proteins associated with RNase E in a multicomponent ribonucleolytic complex. Proc. Natl. Acad. Sci. USA **93:**3865-3869.

Mulligan, M.E. and R. Haselkorn. 1989. Nitrogen fixation (*nif*) genes of the cyanobacterium *Anabaena* species strain PCC 7120: the *nifB-fdxN-nifS-nifU* operon. J. Biol. Chem. **264:**19200-19207.

Murata, N. And H. Wada. 1995. Acyl-lipid desaturases and their importance in the tolerance and acclimatization to cold of cyanobacteria. Biochem. J. **308:**1-8.

Nicol, S.M. and F.V. Fuller-Pace. 1995. The "DEAD box" protein DbpA interacts specifically with the peptidyltransferase center in 23S rRNA. Proc. Natl. Acad. Sci. USA **92:**11681-11685.

Nishi, K., F. Morel-Deville, J.W.B. Hershey, T. Leighton, and J. Schnier. 1988. An eIF-4A-like protein is a suppressor of an *Escherichia coli* mutant defective in 50S ribosomal subunit assembly. Nature **336:**496-498.

Ohmori, H. 1994. Structural analysis of the *rhlE* gene of *Escherichia coli*. Jpn. J. Genet. **69:**1-12.

Pause, A., N. Méthot, and N. Sonenberg. 1993. The HRIGRXXR region of the DEAD box RNA helicase eukaryotic translation initiation factor 4A is required for RNA binding and ATP hydrolysis. Mol. Cell. Biol. **13**:6789-6798.

Pause, A. and N. Sonenberg. 1992. Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. EMBO J. **11**:2643-2654.

Pause, A. and N. Sonenberg. 1993. Helicases and RNA unwinding in translation. Curr. Opin. Struct. Biol. **3**:953-959.

Pharmacia Biotech. 1993. GST gene fusion system. Pharmacia P-L Biochemicals Inc.

Prentki, P. and H.M. Krisch. 1984. In vitro insertional mutagenesis with a selectable DNA fragment. Gene. **29**:303-313.

Py, B., C.F. Higgins, H.M. Krisch, and A.J. Carpousis. 1996. A DEAD-box RNA helicase in the *Escherichia coli* RNA degradosome. Nature **381**:169-172.

Ray, B.K., T.G. Lawson, J.C. Kramer, M.H. Cladaras, J.A. Grifo, R.D. Abramson, W.C. Merrick, and R.E. Thach. 1985. ATP-dependent unwinding of messenger RNA structure by eukaryotic initiation factors. J. Biol. Chem. **260**:7651-7658.

Rippka, R., J. Deruelles, J.B. Waterbury, M. Herdman, and R.Y. Stanier. 1979. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. J. Gen. Microbiol. **111**:1-61.

Rozen, F., I. Edery, K. Meerovitch, T.E. Dever, W.C. Merrick, and N. Sonenberg. 1990. Bidirectional RNA helicase activity of eukaryotic translation initiation factors 4A and 4F. Mol. Cell. Biol **10**:1134-1144.

Rozen, F., J. Pelletier, H. Trachsel, and N. Sonenberg. 1989. A lysine substitution in the

ATP-binding site of eukaryotic initiation factor 4A abrogates nucleotide-binding activity. Mol. Cell. Biol. **9**:4061-4063.

Sachs, A.B. and R.W. Davis. 1990. Translation initiation and ribosome biogenesis: involvement of a putative rRNA helicase and RPL46. Science. **247**:1077-1079.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Sato, N. 1992. Cloning of a low-temperature-induced gene *lti2* from the cyanobacterium *Anabaena variabilis* M3 that is homologous to α-amylases. Plant Mol. Biol. **18**:165-170.

Sato, N. 1994. A cold-regulated cyanobacterial gene cluster encodes RNA-binding protein and ribosomal protein S21. Plant Mol. Biol. **24**:819-823.

Sato, N. 1995. A family of cold-regulated RNA-binding protein genes in the cyanobacterium *Anabeana variabilis* M3. Nucleic Acids Res. **23**:2161-2167.

Sato, N. and N. Murata. 1980. Temperature shift-induced responses in lipids in the blue-green alga, *Anabaena variabilis*: the central role of diacylmonogalactosylglycerol in thermo-adaptation. Biochim. Biophys. Acta. **619**:353-366.

Sato, N. and N. Murata. 1981. Studies on the temperature shift-induced desaturation of fatty acids in monogalactosyl diacylglycerol in the blue-green alga (cyanobacterium), *Anabaena variabilis*. Plant Cell Physiol. **22**:1043-1050.

Sato, N. And N. Murata. 1982. Lipid biosynthesis in the blue-green alga, *Anabaena variabilis*. II. Fatty acids and lipid molecular species. Biochim. Biophys. Acta. **710**:279-289.

Sato, N., N. Murata, Y. Miura, and N. Ueta. 1979. Effect of growth temperature on lipid and fatty acid compositions in the blue-green algae, *Anabaena variabilis* and *Anacystis nidulans*. Biochim. Biophys. Acta. **572:**19-28.

Schmid, S.R. and P. Linder. 1991. Translation initiation factor 4A from *Saccharomyces cerevisiae*: analysis of residues conserved in the D-E-A-D family of RNA helicases. Mol. Cell. Biol. **11:**3463-3471.

Schmid, S.R. and P. Linder. 1992. D-E-A-D protein family of putative RNA helicases. Mol. Microbiol. **6:**283-292.

Smith, L.S., T. Lewis, and S.M. Matsui. 1995. Increasing yield of small DNA fragments purified by silica binding. Biotechniques **18:**970-975.

Stanier, R.Y. and G. Cohen-Bazire. 1977. Phototrophic prokaryotes: the cyanobacteria. Ann. Rev. Microbiol. **31:**225-274.

Stein, J.L., T.L. Marsh, K.Y. Wu, H. Shizuya, and E.F. DeLong. 1996. Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. J. Bacteriol. **178:**591-599.

Thiel, T. 1994. Genetic analysis of cyanobacteria, p. 581-611. *In* D.A. Bryant (ed.), The molecular biology of cyanobacteria. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Thiel, T. and H. Poo. 1989. Transformation of a filamentous cyanobacterium by electroporation. J. Bacteriol. **171:**5743-5746.

Toone, W.M., K. E. Rudd, and J.D. Friesen. 1991. *deaD*, a new *Escherichia coli* gene encoding a presumed ATP-dependent RNA helicase, can suppress a mutation in

rpsB, the gene encoding ribosomal protein S2. J. Bacteriol. **173:**3291-3302.

Wada, H., Z. Gombos, and N. Murata. 1990. Enhancement of chilling tolerance of a cyanobacterium by genetic maipulation of fatty acid desaturation. Nature. **347:**200-203.

Walker, J.E., M. Saraste, M.J. Runswick, and N.J. Gay. 1982. Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1:**945-951.

Wassarman, D.A. and J.A. Steitz. 1991. Alive with DEAD proteins. Nature. **349:**463-464.

Whitton, B.A. and M. Potts. 1982. Marine littoral, p. 515-542. *In* N.G. Carr and B.A. Whitton (ed.), The biology of cyanobacteria. Blackwell Scientific Publications, Oxford.

Wolk, C.P., A. Vonshak, P. Kehoe, and J. Elhai. 1984. Construction of shuttle vectors capable of conjugative transfer from *Escherichia coli* to nitrogen-fixing filamentous cyanobacteria. Proc. Natl. Acad. Sci. USA **81:**1561-1565.

Yamanaka, K., T. Ogura, E.V. Koonin, H. Niki, and S. Hiraga. 1994. Multicopy suppressors, *mssA* and *mssB*, of an *smbA* mutation of *Escherichia coli*. Mol. Gen. Genet. **243:**9-16.

Zhou, C., Y. Yang, and A.Y. Jong. 1990. Mini-prep in ten minutes. Biotechniques **8:**172-173.

Zoller, M.J. and M. Smith. 1987. Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template. Methods

Enzymol. **154**:329-350.

APPENDIX
NUCLEOTIDE AND AMINO ACID SEQUENCES OF gwo29

```
  1 GAATTCGGCTTGTACTAGATGAGGCCGACCGGATGTTGGATATGGGCATTATTCGTGATA 60
      I  R  L  V  L  D  E  A  D  R  M  L  D  M  G  I  I  R  D  I

 61 TCCGTCGTATCCTCTCCCTGTTGCCCAAACAGCGACAAAACTTGCTATTCTTCGCTACAT 120
      R  R  I  L  S  L  L  P  K  Q  R  Q  N  L  L  F  F  A  T  F

121 TCTCAGACAAAATCAAGGAACTCGCCGCCGGTCTACTGAATCGAACGCAGATGATCGAAG 180
      S  D  K  I  K  E  L  A  A  G  L  L  N  R  T  Q  M  I  E  V

181 TAGCACACCGTAACGTTACAGCCGATACTGTGACACAAAAAGTCTACAAAATAGAGCGCG 240
      A  H  R  N  V  T  A  D  T  V  T  Q  K  V  Y  K  I  E  R  D

241 ATAGAAAACGAGATTTACTGGCTCACCTGATTCGGAAAGATAATTGGTATCAAGTACTAG 300
      R  K  R  D  L  L  A  H  L  I  R  K  D  N  W  Y  Q  V  L  V

301 TATTTACTCGTACCAAGTATGGTGCTGACCGTCATGTTAAACAATTAGGCCATGAGCGCA 360
      F  T  R  T  K  Y  G  A  D  R  H  V  K  Q  L  G  H  E  R  I

361 TTCAAGCTCTAGCTATCCACGGTAATAAGAGCCAGTCGGCGCGTACCCACGCTCTGGCAA 420
      Q  A  L  A  I  H  G  N  K  S  Q  S  A  R  T  H  A  L  A  K

421 AGTTCAAAAATGGTAGTTTACAAGTTTTAGTAGCAACTGACATTGCTGCACAGGGACTGG 480
      F  K  N  G  S  L  Q  V  L  V  A  T  D  I  A  A  Q  G  L  D

481 ACATCAGCGAATTGCCTTATGTGGTCAATTTCGATTTGCCCTATGTACCAGAAGATTATG 540
      I  S  E  L  P  Y  V  V  N  F  D  L  P  Y  V  P  E  D  Y  V

541 TTCATCGTATTGGTCGCACTGGTCGCGCTGGTGCATCAGGTGAGGCTGTATCGCTGGTGA 600
      H  R  I  G  R  T  G  R  A  G  A  S  G  E  A  V  S  L  V  S

601 GCGCCGATGAATATCATTTGTTGGCAGATATTGAAAAACTGATTGAAAAGCGATTGCCTT 660
      A  D  E  Y  H  L  L  A  D  I  E  K  L  I  E  K  R  L  P  F

661 TTGAATTGGTAGCGGGTATTGGAGCTAATTCCCAAGCTAAACCCGAACCAACTCAGGATG 720
      E  L  V  A  G  I  G  A  N  S  Q  A  K  P  E  P  T  Q  D  E

721 AACGCAAGCAAAAACCCAAAGATAGTCAGCATCAGCCTCGTCGAGCNATAAGCCGAATTC 780
      R  K  Q  K  P  K  D  S  Q  H  Q  P  R  R  A  I  S  R  I
```